THESIS

LEGIBILITY OF SERIF AND SANS SERIF TYPE FACES IN COMPUTER DISPLAYS

Submitted by

Mark Williams

Department of Technical Journalism

In partial fulfillment of the requirements

for the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Spring 1990

COLORADO STATE UNIVERSITY

December 11, 1989

WE HEREBY RECOMMEND THAT THE THESIS PREPARED UNDER OUR SUPERVISION BY MARK WILLIAMS ENTITLED *LEGIBILITY OF SERIF AND SANS SERIF TYPE FACES IN COMPUTER DISPLAYS* BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE.

Committee on Graduate Work

____

____

____
Advisor

____
Department Head

ABSTRACT OF THESIS

LEGIBILITY OF SERIF AND SANS SERIF TYPE FACES IN COMPUTER DISPLAYS

Studies of type faces printed on paper have generally found minimal legibility differences between the many type faces. Modern computers are capable of displaying many of the same type faces available to printers, but few studies have looked at the legibility of these electronically displayed type faces. This study considers whether type faces with serifs are more legible than those without serifs when displayed on IBM PS/2 8513 monitors, when all other legibility variables are experimentally controlled.

An experiment compared subjects' reading rates for one text set in serif type and one set in sans serif type. Subjects timed themselves as they read two 600-700 word texts from the computer monitors and answered comprehension questions. Variables of subject age, prior use of computers, vision, and use of corrective lenses were addressed.

Analysis of data showed no significant difference in the reading rates of the two type faces. Reading rates for experimental treatments were not significantly different than those for the control, but were significantly different than each other. The variance in the experiment is thus caused by variables other than type face. Uncontrolled variables in experimental design and laboratory set-up appear to have overwhelmed any type face induced effect that may have been present.

Further, better controlled experiments are needed to test the appropriateness of type face for computer displays.

Mark Allen Williams
Technical Journalism Department
Colorado State University
Fort Collins, CO 80523
Spring, 1990

TABLE OF CONTENTS

Introduction

Type faces allow printers to give texts unique appearances, set aesthetic tones, and perhaps most importantly, ease reading by improving letter perception. The average reader, however, is rarely aware of a text's type face, unless it is so poorly designed or reproduced so as to make reading difficult.

Letter perception is a combination of mechanical and cognitive actions. Perceptual psychologists (Sanocki, 1987; Krulee and Novy, 1986) model letter perception as a two-step system. Readers first detect font-specific details: specific shapes, stroke thicknesses, angles, and serif styles. Fonts establish a relationship between characters. All characters in a font act as a set, allowing the eye to perceive groupings of characters (words) more easily. Font-specific details are closely interrelated within each font, producing richly detailed characters economically.

After recognizing font-specific details, readers extract abstract "deep level" information that is true of all letters across most typical fonts, such as the basic letter elements: stroke, loop, bar angle, and dot. At this point readers recognize letters and words for meaning.

The perceptual system for extracting font-specific details tunes itself to the specific font being read, and responds to maximize letter processing efficiency. Mixed-font configurations are therefore read much slower, as they cannot take advantage of font-specific details. Studies (Tinker, 1963; Rehe, 1974) have shown significantly slower reading of mixed fonts, supporting this assertion.

Despite the role type faces play in perception, they are generally chosen based on aesthetics or tradition. These choices are acceptable for normal printed texts, because, as many studies have shown, the type faces in normal use are comparably legible under normal conditions (Tinker, 1963). While script, old English, or novelty fonts are obviously inappropriate for lengthy texts, the standard type faces are quite sufficient for most printed texts.

Type face choices are often considered a triviality to be left to artists, graphic designers or printers. While this may be so for **printed** texts, studies suggest type face is far more important for non-print information displays lacking the resolution, sharpness, and clarity of print on paper.

Indeed, type face choice has been suggested as a prime culprit in the reading difficulty, reader discomfort, and lower reader performance with audiovisual, film, television, and computer displays. These displays require a more careful consideration of type face options.

The ever-expanding use of computer in the workplace and home necessitates a better understanding of the role type face plays in computer displays. Many workers now spend the majority of their days before computer displays, often suffering fatigue, discomfort, and increased stress.

Studies confirm that the reading from CRT computer monitors is harder, slower, and more fatiguing than from paper. While mechanical limitations play a part in this phenomenon, a portion, yet to be accounted for, may be caused by inappropriate type faces. Improving technology now allows computer monitors to display nearly infinite varieties of type faces, where once they were limited to only a few. While CRTs now have the option of varied fonts, they do not yet have the resolution of print on paper. An IBM 8513 monitor, among the higher quality "standard" monitors for personal computers, has a screen resolution only 6.7% of that of commercial printing (IBM, 1987a). Both print and electronic media now share the font options, but they are by no means equivalent.

This study will reconsider type face choice for computer displays. While a many components make up a type face, this study will isolate only a single factor, the presence or absence of serifs, the short cross lines at the ends of the main strokes of letters in most type faces, as its focus. The study asks whether serif or sans serif types are more legible on high resolution CRT displays.

Terminology

Before examining the question of type face legibility in CRT displays, we must establish specific definitions, as many of the terms used in legibility and typography studies have different, and often contradictory, definitions.

Legibility/ Readability

Many studies use the terms *legibility* and *readability* interchangeably to signify ease and speed of reading. The two terms have been defined differently by different disciplines, leading to confusion in comparing studies.

Readability has also been used as a measure of the cognitive difficulty of a text. Texts are ranked by readability and assigned grade levels of difficulty. This confusion between readability as a measure of typography and readability as a measure of text comprehensibility requires that the term not be used in this discussion. This study will use the term legibility exclusively to mean the physical ease of reading.

At the mechanical end, legibility is the "ability to identify a symbol, to the exclusion of all other symbols of a defined set" (Vartebedian, 1971a). Mechanical legibility is a function of visual acuity and is an essential prerequisite for reading (McVey, 1985). Studies of this mechanical legibility do not consider the context of symbol presentation, nor do they attempt to simulate natural reading.

Studies at the opposite end of the legibility spectrum deal with the cognitive factors in "how readers synthesize characters into meaningful units." (McVey, 1985) Cognitive legibility studies consider how words are identified, recognized, and combined into meaningful information.

Most recent studies recognize the dual nature of legibility, and account for both mechanical and cognitive factors to some extent. Military and engineering studies emphasize mechanics because, in many such displays, "message coding is brief, and a single letter may stand alone to represent an entire message. The observer must be able to recognize what that symbol is on the basis of the solitary symbol, with no English context or interpretation" (Shurtleff, 1974).

This study will take a middle ground by recognizing that while legibility is a mechanical measure of vision, all acts of reading involve some level of cognitive processing so cognitive factors must be considered a part of legibility. It will treat legibility as Tinker (1963) treated it. "Legibility deals with the coordination of those typographical factors inherent in letters and other symbols, words, and connected textual material which affect ease and speed of reading." In other words, legibility is, "the ocular and cognitive efficiency in the uptake of information from the display." (Kolers et al, 1981)

CRT

Cathode ray tube (CRT) displays are a broad grouping of electronic displays, including computer monitors, televisions, radar screens, oscilloscopes, and many other displays. They may also be termed visual display terminals (VDT) or visual display units (VDU).

A cathode ray tube is basically an evacuated glass tube with an electron "gun" at one and a screen which is coated with a light emitting material at the other. High electrical voltage produces a stream of electrons from the electron gun. Electronic lenses focus the stream and direct it to the light-emitting coating on the inner surface of the screen. Electrons interact with the coating, causing it to emit light, the light visible to screen viewers. Figure 1 presents a generalized CRT cross-sectional diagram.

All CRT displays use this basic design with variations in beam focusing, beam movement, distortion, and screen components. Computer monitors, the focus of this study, are among the most sophisticated CRTs produced (Cakir et al, 1980).

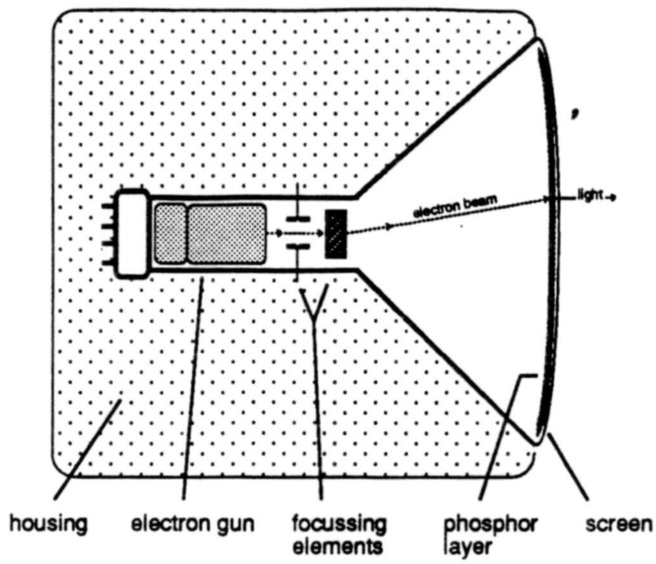

Figure 1
Basic CRT construction (cross section). Adapted from Cakir et al. (1980)

The density of pixels, the individual light emitting elements which make up the CRT screen, determines the unit's visual quality, display letter sharpness, and what letter features can be displayed. Resolution, a measure of pixel density and screen quality, varies with CRT manufacturers and models, and has improved steadily with the development of new technologies. Currently manufactured IBM 8513 displays feature resolutions of 3.40 pixels per mm, or roughly 86 pixels per inch (IBM, 1987a). Early CRT displays lacked the technology for high pixel density and had resolutions as low as 1.8 pixels per mm (Shurtleff, 1980).

This study considers high resolution CRTs as those with densities of 3.0 pixels per mm or greater. This is a generous grouping by industry standards, but it is appropriate for relating this to previous studies.

Type Face/Font

Both *type face* and *font* refer to the various styles of type available for print. Thousands of type faces exist, each with its own combination of geometry, height/width ratio, stroke width, weight, and style (Baskette et al, 1986).

Serif and *sans serif* are the two type groupings to be considered in this study. They are distinguished by the presence or absence of serifs, the short cross lines at the ends of the main strokes of letters in most type faces. Serifs may be pointed as in "oldstyle" types, squared off in "modern" types, or square or blocked in "square serif" types (International Paper, 1988). Figure 2 illustrates the differences between the two type groupings by pointing out serifs.

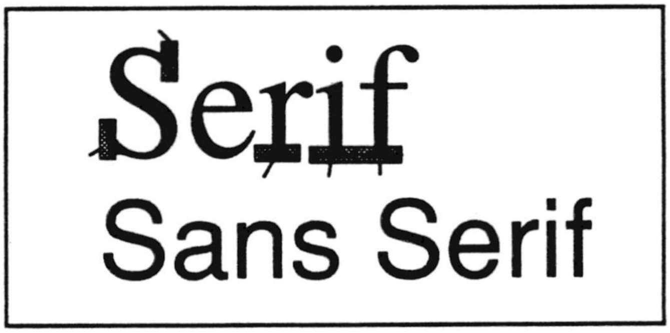

Figure 2
Serif and sans serif types

The Legibility Variables

Many equally relevant and often confounding variables cloud discussions of any element of legibility. In any reading situation, legibility is the product of the reader, the material, the way in which material is presented and the reading environment. No primary and secondary variables exist; all are equally relevant and interact in most reading situations (Gould et al, 1987; Tinker, 1963). Paterson and Tinker (1940a) report that no single variable has more than about a 10% effect on reading rate. Isolating individual variables for study is thus an unnatural although necessary task.

Classical Variables

The classical variables have received much attention, from as early as the late 1800s. These variables include *character size, line width, interline spacing, case, color, contrast, illumination, and type face.*

Character size

An imprecise measurement system, and differences between absolute and perceptual size, complicate discussions of character size. Type is measured by points (1/72 inch), but this system is not standardized between type faces. This lack of standardization stems from early "hot type" printing in which individual characters were cut into small metal blocks. These letter blocks were assembled by hand into lines of type, which were composed into pages and printed. The "point" measurement of a letter was the measure of the metal block upon which the letter was cut, rather than the measure of the letter itself (Rehe, 1974; International Paper, 1988). Even though type blocks are no longer used, the imprecise "point size" measurement system still predominates modern typography.

While point size measurement can be used to compare type sizes within the same type face, the system is inaccurate for comparisons between different type faces. The metal blocks of a Times Roman 12 point "h" and a Helvetica 12 point "h" are the same size: 12 points. The letters, however, are slightly different sizes because of their different designs. A more precise measurement system, the "x-height" measure (the vertical measure of the "x" of the type face) is necessary for comparisons between different type faces.

While "x-height" can standardize absolute sizes of print on the display surface, optical factors also play a part in character size. Readers perceive print displayed at one foot away to be much larger than the same print at three feet. To account for this optical factor, human factors psychologists measure type not by its absolute size on the display surface, but by the angle it enters the eye and focuses on the retina. Thus, a character with an x-height of 2.1mm displayed at 50cm would measure and angle of 14' of an arc. The same character at 100cm measures 7' of an arc. Conversely, a 2.1mm character viewed at 50 cm has the same optical size as a 4.2mm character viewed at 100cm (Cakir et al, 1980).

The relative measure of size recognizes human optical perception, and is particularly useful in selecting an optimum size for a fixed position reading environment such as a theater or airplane cockpit. In most situations, however, readers can choose the distance from the display, so visual angle measurement is not as relevant.

Standard legibility guidelines for printed characters on paper suggest an optimum character size of 8-10 points (Felkner, 1981) or 9-12 point (Tinker, 1963). Optical size should be 0.035 inches per foot of viewing distance (Murrell, 1965), a value which equals roughly 10 minutes of an arc. Guidelines for electronic displays suggest a slightly larger character size of 2-5 mm (Smith, 1979) or a visual angle of 12-15 minutes of an arc (Gould, 1968).

Line width

Line width affects how well a reader can move through a text. Lines that are too short prohibit the eye from making maximum use of horizontal perceptual cues. Lines that are too long cause inaccuracies and difficulties in relocating the beginning of each new line. (Paterson and Tinker, 1942).

For the optimal type sizes, 10 to 12 words per line, or lines of about 3-4 inches long are most comfortable for the eye (Rehe, 1974). Tinker and Paterson determined the optimum width to be 80 millimeters, just over 3 inches. Other studies (Paterson and Tinker, 1929; Paterson and Tinker, 1940b; Paterson and Tinker, 1942; Tinker, 1963; Felkner, 1981) report a line width somewhere between 3 and 4 inches seems to be best for 10 point type.

Interline spacing (leading)

Spacing between lines of type can have a significant impact on how easily the type is read. Interline spacing, or *leading,* prevents lines from blurring, and increases reading speed. Material with no spacing between lines is read relatively slowly, and reading rate increases with additional leading (Bently, 1921). Paterson and Tinker (1932) found that optimum leading can increase reading speed up to 8% over text with no leading. They suggest leading of 1 to 4 points for 10 point type on a 3 inch line. Optimum leading depends on the type face used, as heavier types require more leading. Gould (1968) suggests a slightly higher leading, 30-50% of character size, be used for CRT displays to prevent blurring.

Case

Words are perceived by their specific word-shape outline, which is unique for lower-case words. Words set in all-caps do not provide specific word-shape outlines since they produce an oblong, uniform word-shape, so readers can't benefit from the distinct features of ascenders and descenders, as illustrted in Figure 6. The eye cannot perceive word-form cues, so it must decipher the word, letter by letter (Rehe, 1974). Tinker and Paterson (1928) found that subjects reading type set in all-capitals averaged a rate of 4.74 words per second, while those reading type set in lowercase averaged 5.38 words per second.

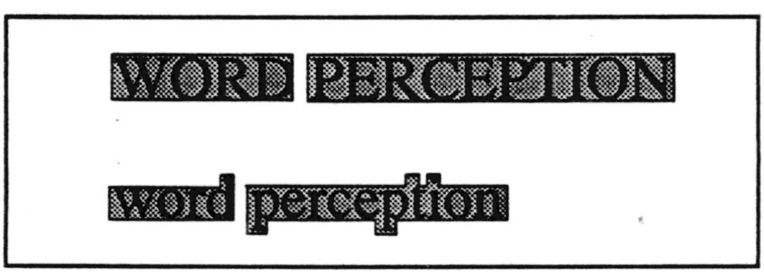

Figure 3
Word form cues for upper case and lower case types. Adapted from Rehe (1974).

Color and Contrast

Print on paper is visible because dark print absorbs light, while light paper reflects light. The *brightness contrast,* the perceived blackness of type against the brightness of paper background, is an important determinant of legibility (Tinker and Paterson, 1931). Numerous studies examining combinations of colored ink and paper find that the most important factor seems to be the brightness contrast between colors of print and paper. Studies by Hackman and Tinker (1957) and Tinker and Paterson, (1931) found that black ink on white paper and black ink on yellow paper were most legible.

Reversed print, or white letters on dark background, is frequently used for emphasis, although such print is significantly less legible. Holmes (1931) found that reading of 10 point reversed type was 14.7% slower than normal type. Paterson and Tinker (1931) found a 10.5% decrease in speed.

Tinker (1963) suggests, "when white type on black is used, the amount of text should be small, and a sans serif type should be used to minimize loss of legibility which ordinarily occurs with such arrangements."

Illumination

Reading requires ample lighting or illumination so that characters can be distinguished from their background. Moderate intensities of 20-30 foot-candles are adequate for most reading. Reading small print or materials with low brightness contrast between print and paper requires a brighter light. Elderly people or those with less than normal visual acuity require brighter light. Few reading situations require intensities of over 50 foot-candles of light. (Tinker, 1963)

Type face

"Which type face is best?" has been debated nearly as long as varied fonts have been available to printers. Pyke's (1926) comprehensive survey of 183 legibility studies prior to 1920 found little validity or reliability in studies. Most conclusions were based, at best, on weak research, and, at worst, on pure opinion.

Scientific studies have addressed the question of optimum type by experimentally testing the legibility of type faces. Studies by Roethlein (1912), Paterson and Tinker (1932), Burtt and Basch (1923), Webster and Tinker (1935), Luckeish and Moss (1937), Tinker and Paterson (1942, 1943), English (1944), Coleman and Kim (1961), Poulton (1965), Pittman (1976), and Jha and Daftuar (1981) found minor variations between fonts, but concluded, as had Roethlein, that "type faces in common use do not differ greatly with respect to the speed with which they are read."

Paterson and Tinker (1940a) propose that the nearly equal legibility of commonly used type faces is a product of a typographic evolution. "Such (commonly used) type faces have survived in competition with hundreds or even thousands of type faces designed and put on the market. Those type faces that survive the relentless competitive struggle for existence are shown by our experiment to be equally fit."

Only one sans serif type face, Kabel Light, was included in the previously mentioned studies, primarily because except in advertising, sans serif type was rarely used prior to the 1940's (Hvistendahl and Kahl, 1975). Kabel Light proved comparable to the other fonts for legibility, and Paterson and Tinker (1932) suggest that "Kabel Light might prove to be slightly more legible than the other type faces ... were our readers accustomed to it."

This assertion that legibility of type faces may be merely a function of familiarity warrants further investigation. At the time of the Paterson and Tinker study, few readers had read sans serif text. Modern readers have likely read significant amounts of sans serif text, and would therefore be much more familiar with such types. A modern replication of the study might test Paterson and Tinker's familiarity assumptions.

Several psychologists have asserted that serifs play a significant role in letter differentiation and thus legibility. Burt (1953) claims that serifs make type faces more readable because they emphasize the distinguishing features of letters. Cheetham and

Grimbly (1964) state, "if the characters of a type face have too much of a family resemblance, its words will also tend to look inconveniently alike."

Robinson et al (1971) suggest that, "the neurological structure of the human visual system benefits from serifs in the preservation of the main features of letters." Because letters are distinguished by their outline features, serifs add more unique strokes to ease distinction. This is a particularly important feature with small print. Dowding (1959) suggested that sans serif types are harder to read and take 7.5% longer to read than serif types.

Hvistendahl and Kahl (1975) tested reading rates and reader preferences for newspaper texts presented in serif and sans serif type. Subjects read the serif text 7-10 words per minute faster than the sans serif text, and almost two thirds of subjects preferred serif type to sans serif when given a choice. Data suggested that younger readers (15-35) and light readers tended to read the two type faces at about the same rate, while heavy readers read serif faces faster.

McVey's 1985 study of typography of audiovisual materials found markedly different results. "Although quite legible as type face in a book, serif symbols and letters having a variable stroke width are not recommended for TV or film projection display. The thinner stroke elements that make up part of a symbol's design frequently are lost either in making the visual or displaying it."

Electronic Display Legibility Variables

Print legibility research has become rather static in the past 25 years. Printing processes have remained relatively unchanged, at least in terms of appearance of final product. Most legibility issues have been studied and concluded. In the same time period, expanded computer technologies have brought electronic displays into the daily lives of the majority of readers. Electronic information displays are now replacing print media, and bring with them a whole new realm of legibility issues.

Electronic displays share many of the print legibility variables, but also have additional variables of their own. Legibility guidelines developed for printed texts do not, therefore, directly apply to these displays. The overwhelming growth of computers in the workplace and home necessitate a reconsidering of legibility findings.

Many of the classical legibility factors such as character size, interline spacing, case, color, contrast, line width, and type face directly apply to text displayed on CRTs. Electronic displays also have additional legibility variables, including CRT physical dimensions, presentation method, luminance, irradiation, glare, color, and element burnout.

CRT dimensions

CRT displays have physical dimensions much different than those of most printed materials. Printed texts normally have proportions of 1.414: 1 (height/ width ratio). CRT displays, however, are horizontal rather than vertical, with proportions of 1:1.414. Furthermore, CRT screens can display far less text than an equivalent printed page - up to 50% less. A standard printed 8.5" x 11" page will therefore require up to two full CRT screens (Rubins and Krull, 1985).

Standard CRT dimensions are inefficient and unnatural for readers and therefore may slow reading. Kruk and Muter (1984) found that reducing printed text to the dimensions of standard CRT displays reduced reading speed by 10%.

Presentation method

CRTs display characters by a matrix of dots related to the screen density. Character must be of at least a 5x7 matrix (Gould, 1968), although 7x9 or higher matrices improve legibility (Maddox et al, 1977; Shurtleff, 1980). Circular dots are superior to elongated dots (Vartebedian, 1971a). "The more dots available, the better the resolution of symbols" (Shurtleff, 1980). Improved CRT technology now allows displays using 7x15, 9x16, and 12x20 matrices (IBM, 1987a; IBM, 1987b; IBM, 1987c; IBM, 1987d).

Light

Illuminance, or the quantity of light striking a surface, is a critical variable for legibility of printed texts. *Luminance*, the quantity of light emitted, is its correlate for CRTs. Luminance is a crucial component of CRT legibility because CRT characters are visible due to the contrast between light-emitting elements and non-light-emitting elements of the screen.

Quantifying luminance on CRTs is quite difficult, as the light-emitting pixels are very small, the light is intermittent, and the amount of available light may vary between different regions of the screen (Cakir et al, 1980). Light emitted from the CRT phosphor breaks down as it exits the screen surface. The distinct beam of light breaks and spreads as it leaves the screen, creating blurring, or *irradiation.*

CRT-produced light must compete with light striking the screen from the outside. As shown in Figure 4, part of this external light, from office fixtures or windows, reflects off the screen, producing *glare.* Another part penetrates the screen surface and lightens the display background, lowering the luminance contrast between character and background (Cakir et al, 1980).

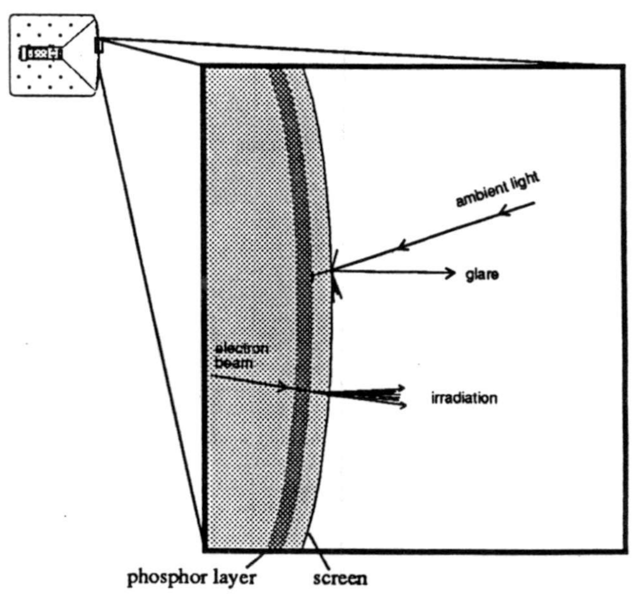

Figure 4
Lighting variables in CRT displays. Adapted from Cakir et al (1980).

Color

Traditional CRT displays presented colored, light-emitting characters on a dark background. Phosphor elements in screen pixels emit a colored light, ranging from white to yellow green, to green, to violet, depending upon the chemical makeup of the element. The normal light-adapted eye is most sensitive to light in the green part of the spectrum. The majority of monitors therefore display green or yellow-green characters (Cakir et al, 1980).

As with print, color of characters is far less important than the relationship between color of character and color of background. Studies comparing character/background color combinations have found that contrast between the two colors is crucial for legibility (Nilsson et al, 1983; Carter, 1982; Carter and Cahill, 1979; Christ, 1975).

Flicker

Light emitting elements in CRT screens must be continually stimulated by the electron gun to emit "constant" light. Early CRTs could not re-stimulate the elements fast enough, so characters appeared to flicker. Technological improvements have eliminated visible flicker by increasing "refresh rates" above that the threshold of human visibility. Modern computers feature rates as high as 100 cycles/second, nearly double the rate necessary to eliminate visible flicker (Cakir et al, 1980).

While flicker is not visible, it nevertheless exists in all computers. Display pixels continually pulse as the electron beam strikes them. Wilkens (1986) claims that this undetectable flicker interrupts ocular motor control of eye movements and therefore slows reading from CRTs. His eye-movement experiments comparing CRT displays, print lighted by intermittent fluorescent light flickering at the same frequency as the CRT, and print lighted by constant (non-flickering) light support this claim. CRT and flickering light prints were read very similarly, both significantly slower than constant light print.

Element burnout

As CRT screens age, individual pixels may burnout or misfire. Riley and Barbato (1983) studied legibility of type faces under simulated pixel burnout. While certain characters were more vulnerable to burnout-induced misperception, none of the fonts tested were any more resistant to burnout-induced errors.

Typography

Even high resolution CRTs (80+ pixels/inch) are far from the resolution standards of laser print (300 dots/inch) or commercial typesetting (1270 to 2540 dots per inch) (Fenton, 1989). Early electronic displays lacked the technology of today's screens and had resolutions as low as 32 pixel/inch (Shurtleff, 1980).

Most early studies set out to find the optimum type face for electronic displays based on studies of character height, stroke width, height/width ratios, and type face geometry (Atkinson, 1952; Brown, 1953; Huddleston, 1974; Maddox et al, 1977; Shurtleff and Owen, 1966; Shurtleff et al, 1966). These studies did not compare alternative type faces, but rather set out to redesign a face suitable for electronic use. They suggest that alterations such as increasing the size of small distinguishing details, changing the overall configuration of symbols, and altering the curvature of symbol strokes can improve legibility of individual characters (Shurtleff, 1980).

Roland and Cornog (1968) studied user opinions about the alphanumeric characters displayed on the FAA's air traffic control radar systems. They concluded that none of the available fonts were adequate for the quick recognition tasks of air controllers, and developed *Courtney*, a new "ideal" electronic display type face.

Shurtleff and Owen (1966), and Shurtleff et al (1966) refuted the conclusions of Roland and Cornog by conducting short exposure experiments rather than simply recording users' opinions about legibility. They demonstrated that with several minutes of practice, subjects could identify characters of all commercially available fonts similarly.

Early work on computer display legibility by Gould (1968) and Vartebedian (1971a, 1971b) used short exposure methods to set legibility guidelines for CRT displays much as Tinker (1963) had done for paper text. These and most subsequent studies used short-exposure recognition methods to assess legibility.

Recent studies present broad generalities about type faces for CRT use. No standard style has dominated displays, and most research concludes only that, "type must have uniform intensity and illumination, and must emphasize differences in salient feature: roundness, straightness, angularity" (Rubins and Krull, 1985).

Methods of Investigating Legibility

Because of the many variables involved in both print and electronic legibility, methods of investigating it are varied and diverse. Legibility experiments have historically fallen into four categories: distance methods, distortion methods, eye movement methods, and time/rate methods. All four types of experiments have validity for assessing certain components of legibility, although none are applicable for all studies.

Distance methods test the boundaries at which text can be perceived. They provide a measure with which to compare different types. *Perceptibility at a distance* measures the distance from the eyes at which printed symbols can be perceived accurately. *Perceptibility in peripheral vision* , a variation of perceptibility at a distance, measures the horizontal distance from a fixation point at which a printed symbol can be perceived accurately.

Distortion methods distort or break down text to see what types are most resistant and therefore most legible. The *Visibility* method uses the Luckeish-Moss Visibility Meter, an apparatus which distorts subjects' vision to determine the threshold visibility of printed material (Luckeish and Moss, 1935). The *Focal Variator* method, artificially distorts characters projected on a screen. Both methods yield a measurement of minimum legibility and allow comparisons between type faces.

Eye movement methods monitor subjects' eye movements during reading to determine legibility. The *Reflex Blink Technique* measures the subject's rate of blinking of during reading, assuming that any factor which reduces the ease of seeing will increase the frequency of blinking. Studies comparing this technique to other proven methods found low correlation and low validity (Tinker 1945; Tinker 1947; Tinker 1948).

The *Eye Movements* technique monitors eye movements during "normal" reading, using a corneal reflection apparatus (Tinker, 1931) or computer controlled eye monitor (Kolers et al. 1981). This method generally yields valid, reliable legibility measures, and can give information on why non-optimal typography is read more slowly than optimal typography (Tinker, 1936). The apparatus detects the location of legibility problems in a text, and is therefore a very descriptive evaluative tool.

Time/rate methods time the act of reading or perception to assess relative speed of reading or perception of different type faces. The *Speed of Perception* (short exposure) technique measures the quickness and accuracy with which letters, digits, words and phrases can be perceived, and ascertains the "recognizability" of printed symbols. The short exposure method is used extensively for CRT legibility studies and is almost the exclusive method for assessing the legibility of electronic systems (Shurtleff, 1980).

The *Rate of Work* technique has been used widely since 1896. Variations include speed of reading, amount of reading completed in a set time limit, time taken to find a telephone number, time taken to look up a power or root in mathematical tables, and work output in a variety of situations which involve visual discrimination. "The method has come to the accepted as the most valid technique for studying the legibility of printed material" (Tinker, 1963).

The rate of work test of legibility simulates natural reading. Tinker (1963) states, "Other things being equal, a typography that is read faster than another should be easier to read. When one arrangement is read significantly faster, legibility must be a factor of importance."

Comparing the methods

Tinker (1944) analyzed visibility, perceptibility, and speed of reading legibility measurements for a number of heavily-researched type faces to investigate correlations between the methods. Visibility and perceptibility measurements had a high correlation with each other, but low correlation with speed of reading. The lack of correlation suggests that the three tests cannot be compared interchangeably. According to Tinker, speed of reading is the most valid test of legibility of continuous text because it most closely simulates natural reading. Visibility and perceptibility represent abnormal, artificial reading situations.

Visibility and perceptibility scores can also produce absurd conclusions. For instance, large type sizes (18, 24, 36 pt.) yield high visibility scores, but they produce print which prevents formation of effective word-form clues and use of peripheral vision—two major components of reading. Furthermore, a text printed in large type would be massive, inefficient and expensive.

Current legibility research

Modern technology has brought the computer to the majority of business and offices. Many workers now spend the majority of their work hours before a CRT, often doing prolonged reading rather than simply identifying symbols. The intensity of this user/machine interaction necessitates further examination of the legibility of computer displays.

While short exposure experiments were valid for military and aviation displays, they do not closely simulate normal, day-to-day computer use. Results from early CRT studies may therefore lack validity for answering modern CRT legibility questions. Furthermore, improved technology has introduced new parameters, such as varied type faces, to CRT legibility considerations.

As computer and other electronic displays have grown, so too has the quality of the display screen. Modern computers have moved to high resolution screens capable of displaying black characters on white backgrounds, much like print on paper. Furthermore, these new displays are capable of using varied type fonts and sizes, and formats. As these displays move closer to traditional print, we must consider whether "print" or "electronic" legibility guidelines are more appropriate for high resolution displays.

Studies have consistently documented visual fatigue (Mourant et al, 1981; Matula, 1981) and 20-30% slower rates of work (Gould and Grischkowski, 1984; Muter et al,1982; Duchnicky and Kolers, 1983; Kruk and Muter,1984; Wright and Lickorish, 1983; Gould et al, 1987) for subjects using CRT displays rather than paper.

Kruk and Muter (1984) blamed approximately half of the decreased speed on screen formats that could display only short line lengths and sub-optimum leading between lines. They could not account for the remaining decrease. Gould et al (1987) found that the reading speed difference between paper and CRT is about the same magnitude as the reading speed difference between "good-quality print on paper and absolutely unacceptable-quality print on paper." Creed et al (1987) suggest that "character font may be a major factor in the poor performance with the CRT."

The growth of CRT type face options has not been followed by research evaluating these faces. CRTs still retain many of their special legibility variables, yet type face decisions are presently made on data from studies of printed text. These data are not sufficient for the CRT context, particularly with the extensive, heavy, and still growing daily use of computers.

While some computer programs allow user selection of display type face, size, and format, most non-text-related programs (spreadsheets, databases, etc.) offer little or no user control of how the information is displayed. On-line materials, command ("dialog") displays, and other non-interactive materials likewise offer little user control. In effect, programmers control the majority of information displayed on computers. They need to know what type to use for optimum user efficiency and comfort.

It may be argued that how information is displayed is a minor, even trivial factor in computer user problems. In comparison to reducing glare or modifying work style, improvements in information presentation will only slightly improve work speed and comfort. When considered in the context of the many hours spent by workers in front of CRTs, even minor improvement can save thousands of hours and millions of dollars.

In a study of information formats used by Bell Systems computers, Tullis (1983) compared alternative display modes to improve the company's CRT displays. While his improved screen format reduced reading time by only one second, the net time savings for the company in one year amounted to 55 person-years, or roughly $1,000,000.

Creed et al (1987), Gould et al (1987), and others suggest that type face is a major factor in poor reading performance with CRTs. No studies in the public literature have fully addressed the issue of CRT type face in the light of modern CRTs capable of displaying complex, varied types. Until CRT type face issues are addressed, programmers will continue to select display type faces based on convention, taste, programming efficiency, or guidelines for printed text.

Research Question

This study examines whether some type faces are better than others for use in high resolution CRTs. Failure to identify and use more legible type faces could waste huge sums of money, consume millions of hours of worker time, and cause unnecessary user fatigue.

This study compares the legibility of representative serif and sans serif type faces displayed on an IBM PS/2 8513 color monitor. In this study, legibility is a measure of visual and cognitive ease of reading, as measured by reading rate of continuous text read for comprehension. While legibility is a combination of a number of factors, this study isolates only those directly attributed to type face: type geometry, character weight, stroke width, style, and presence of serifs. This study further limits these legibility variables by comparing two types that are as close as possible in geometry, character weight, stroke width, and style. While some variation exists due to the very nature of the commercial type faces, the most significant difference between the faces studied is the presence of serifs. Thus this study measures the legibility differences caused by the presence or lack of serifs on letters.

This study asks, "Given controls for the legibility variables not attributed to type face, is there a significant difference in an individual college reader's reading speed of a 650-700 word text set in 9 point Helvetica Medium (sans serif) type and a comparable text set in 10 point Times Roman (serif) type when displayed on an IBM PS/2 8513 color CRT monitor?"

Methods

This study employed a variation of the rate of work procedure used most frequently and successfully in comparable legibility studies. While rate of work methods are considered valid measures of legibility, they pose special problems of administration. First, individuals have different reading rates. We can not, therefore, simply give one group of subjects a text set in one type face and give a second group of subjects the same text set in a different type face, and hope to obtain meaningful results. The range of reading rates is so large that variation in performance caused by type face would not be detectable. Rate of work experiments must therefore compare each individual's performance on one treatment against the same individual's performance on another treatment in order to measure variances caused by legibility. (Tinker, 1963)

This poses a second problem: developing two separate reading texts that are similar in reading difficulty but different in content. To compare two treatments, subjects must read two separate texts. The texts must be distinctly different, as subjects will read a text differently if they have read it once before and already know its content. Texts are so complex that developing comparable readings is quite difficult. For valid measure of effects, both readings must be comparable in vocabulary, sentence length, topic difficulty, grammatical structure, and format.

While every effort can be made to develop two very similar texts, individual subjects assess these many factors differently. Two texts that are very similar to some subjects may be quite different to others. Measures of the reading rate difference between the two may therefore reflect differences in texts rather than differences in legibility.

This experiment tried to overcome these problems, as well as follow the guidelines posed by Tinker (1963) to assure validity. Tinker claimed that rate of work tests must include:

-reading material uncomplicated by comprehension difficulties

-comparisons between reading materials of equal difficulty

-sufficient reading material and sample size

-adequate checks on comprehension

For this experiment all subjects read two texts set in two different type faces from a controlled CRT display. The experiment measured the time needed for subjects to complete the texts, and compared each subject's reading rates for the serif and sans serif texts. This method, measuring completion time of a given amount of text, is equally valid as measuring volume of reading completed in a given time (Tinker, 1963). Comprehension questions following each text assured that subjects read for comprehension rather than speed.

The Laboratory

The study used the Colorado State University Department of Technical Journalism Publishing Laboratory for all experiments. A network of 20 IBM PS/2 Model 60 personal computers served as apparatus for the experiment. Each computer ran an IBM PS/2 Color Display 8513 monitor, featuring a 12 inch (8.38"W x 6.13"H) screen. Displays had a high screen refresh rate (no flicker) and resolutions of 85.9 pixels per inch (IBM, 1987a). All computers and monitors had been used in the lab for 18 months, and had no visible malfunctions.

Treatments for this study were prepared using Aldus Pagemaker software supported by Microsoft Windows. This application displays black characters on a light background, and is capable of displaying many different type faces in 6 point to 72 point sizes (Aldus, 1987; Microsoft, 1987). Pre-tests using students in the laboratory found that little variation between serif and sans serif type faces is evident with 6 point type, but 8 point and larger serif and sans serif type faces can clearly be distinguished from each other. Serifs were clearly evident in the displays, although varied stroke width, a normal feature of printed serif types, was noticable only with very large type sizes (24 point and greater).

Lighting

Lights in the laboratory were turned off during the experiment to eliminate confounding ambient light effects. The small amount of ambient light present in the lab originated from the CRTs themselves. The etched glass screen in the monitors minimized the remaining glare.

The CRTs luminance was controlled by manually setting display brightness via monitor controls. For the simple standardization necessary for this experiment, all monitors were set for maximum brightness. Subjects could not change monitor settings during the experiment. Display contrast was likewise controlled by maximizing the user-selected contrast. While lack of contrast has been shown to significantly slow reading (Tinker, 1963), excess contrast has not shown the same effect. Physical photometer measurement showed consistant luminance throughout the reading portion of the screen.

In normal use, CRT users would select the most comfortable combination of brightness and contrast for their tastes. The dials used for such settings have no increments, so standardization between the 20 monitors used for this experiment required that settings be at the highest value (end of dial).

Texts

Each subject read two separate texts, one set in serif type and one in sans serif type (experimental treatment) or both set in sans serif type (control). The experiment used texts extracted from passages in the Nelson-Denny Reading Test (Brown et al, 1981), a source of reading material for several other legibility studies. The passages were designed to evaluate reading rate of high school readers. Test statistics and passages used for the experiment are contained in Appendix A.

Passages were 650-700 words each, designed to be read by an average reader in 3 to 5 minutes. This rather brief reading period is sufficient for assessing legibility for any length text, as, "any effect of typographical factors upon speed of reading found for brief work periods of 1.5 or 1.75 minutes hold for longer periods up to 30 minutes" (Tinker, 1963).

Comprehension Test

Comprehension questions developed for the passages from the Nelson-Denny Reading Test measured text comprehension. Subjects answered four multiple-choice (five choice) questions for each treatment.

Readers generally vary their rate of reading to maintain a constant level of comprehension. That is, they read harder texts slower than easy texts to gain a similar level of comprehension (Tinker, 1963). The experiment set out to measure speed of reading under natural reading conditions (reading for comprehension). Because the speed of reading was the focus of the study, some subjects may have read as fast as possible, sacrificing comprehension.

Subjects

Study subjects were students in an introductory news writing class at Colorado State University. Sixty-nine students from four classes volunteered to participate in the study, understanding that participation in no way affected their grade in the class. Subject ages ranged from 18 to 49, with a mean age of 24.5. At the time of administration of the study, subjects had used the IBM PS/2 computer system for class work for 7 weeks, and were familiar with basic operations, including use of the mouse to scroll down the screen.

Visual acuity is essential for efficient reading, so subjects reported whether they could read the experimental text without straining, if they were wearing prescription glasses or contact lenses, and how many hours they had worked with a computer on the day of the experiment (an indicator of visual fatigue).

Four subjects reported that they could not read the text without straining and were removed from the data pool. Fifteen subjects reported wearing prescription glasses, and 17 reported wearing contact lenses. Ten subjects had worked with a computer for an hour or more prior to the experiment. Four subjects reported errors in computer use such as mistakes in scrolling which made the text disappear from the screen. Such errors contaminated reading rate measures, and were therefore removed from the data pool.

These exclusions left 56 "clean" scores for analysis.

Text Treatment

Text displayed on the screens was set in 3.5 inch (108 mm) width lines with 3 points leading - an optimum arrangement by the guidelines of Rehe (1974), Tinker (1963), and Gould (1968). The texts filled roughly 6.5 screens. Subjects scrolled through the screens using the mouse, as they were accustomed to doing with the Microsoft Write program that they have used for class work.

The experiment used five combinations of the two texts and two type faces. Serif treatments were set in 10 point Times Roman type, a "modern" type style commonly used in newspapers, featuring thin, squared-off serifs. Sans serif treatments were set in 9 point Helvetica Medium, a sans serif type even in overall weight with very little contrast between thick and thin strokes (International Paper, 1988).

The Times Roman and Helvetica treatments had different point sizes (10 and 9, respectively), but were the closest possible match in letter size. The "X-height" on the screen was physically measured as 1.8 mm for Times Roman and 1.8 mm for Helvetica. While the two treatments were slightly different in size, they were the closest available types appropriate for this study. Figure 5 illustrates the difference between the two type faces. At a distance of 50 cm (an average arm's length from the screen) characters measured 12' of an arc, well above suggested minimums.

This is 9 point Helvetica Medium.	The quick brown fox jumped over the lazy dog.
This is 10 point Times Roman.	The quick brown fox jumped over the lazy dog.

Figure 5
Examples of the two type faces used in experiment.

Order of presentation of texts and type faces was varied to control for familiarity effects and minor differences between texts. A random numbers table was used to randomly assign treatments to computers within the laboratory to prevent bias between different classes, and between different seats within each class. A control treatment (treatment 5) presented both texts in sans serif type to test the similarity of text A and text B. These treatments were developed as validity check to control for order of presentation effects, not as a primary focus of the experiment. Table 1 lists the presentation orders of each experimental treatment. Figure 6 illustrates the random assignment of treatments to computers within the laboratory.

Distance from screen

Subjects sat at an arm's length from the screen. This crude measure of distance assured that subjects were well within suggested perceptual accommodation ranges (Cakir et al, 1980). More accurate standardization of perceptual distance such as physically positioning subjects' heads in a brace or other device would be favorable for this experiment, but was not feasible in the laboratory setting.

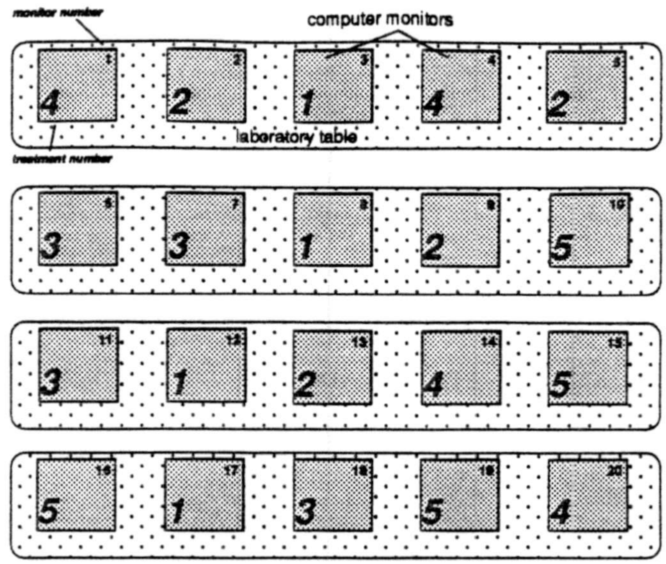

Figure 6
Random assignment of treatments to computers in the laboratory.

Table 1
Presentation orders of experimental treatments

Treatment	1st reading		2nd reading		*n*
	font, text		*font, text*		
1	serif	A	sans	B	11
2	sans	A	serif	B	12
3	serif	B	sans	A	10
4	sans	B	serif	A	10
5	sans	A	sans	B	13

Scrolling

This study required subjects to scroll text down the screen, as a single screen could hold only a sixth of the experimental text. All subjects were familiar with scrolling on the PS/2, and had used the scrolling feature regularly in their work for 7 weeks.

Scrolling adds a confounding factor to the experiment, although Chen et al (1988), Chen and Tsoi (1988), Granaas et al (1984), and Duchnicky and Kolers (1983) have shown that scrolling method had minimal effects on reading speed. The subjects' familiarity with scrolling on the PS/2 machines should prevent scrolling from slowing reading times.

The experiment administrator asked subjects to scroll through the text line-by-line and conducted an exercise to allow subjects to practice line-by-line scrolling. Despite this exercise, some uncontrollable variations may have occurred in the method of scrolling, because some subjects may have read a number of lines, then scrolled to the next full screen of fresh text, read the entire screen, and so on. We can assume, however, that subjects used the same scrolling method for text 1 as they did for text 2. Reading rates were compared only between these two treatments, not between subjects. Scrolling was not, therefore, a major factor in reading rate measure.

Procedure

The administrator pre-set the experimental displays prior to the beginning of each class. Each display featured a 199mm x 81 mm text area and 47mm x 49mm timer. An introductory screen was displayed on the monitors prior to the beginning of the experiment. Figure 7 illustrates the experimental display as seen by subjects.

The administrator introduced and discussed the project, emphasized that participation would not influence subjects' grades, and gave subjects the option to decline participation. Subjects read and signed a subject information/consent form.

Subjects began by completing a data form that asked their age, whether they wore glasses or contact lenses, if they had difficulty reading the words displayed on the mask, and how many hours they had worked on a computer on the day of the experiment. The administrator again quickly reviewed the test procedure. He emphasized that subjects should read naturally, but for comprehension, as a comprehension test would follow each

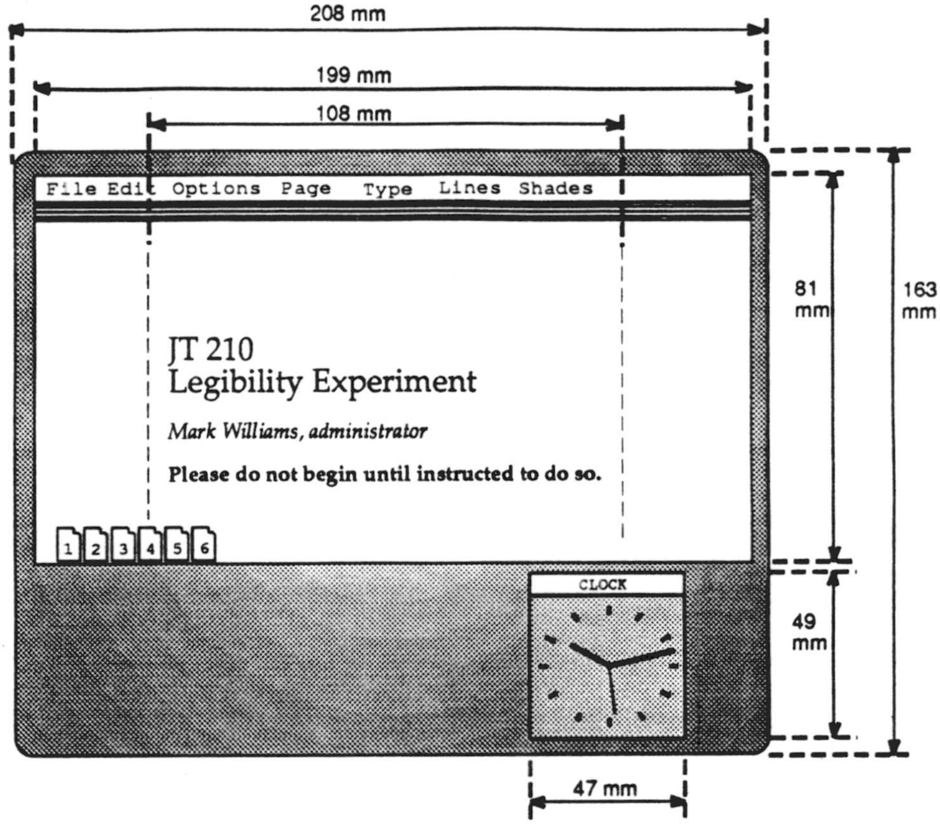

Figure 7
Experimental display as seen by subjects.

passage. Emphasizing comprehension assured subjects read as they would when normally extracting information.

Next, the administrator explained how to scroll down the experiment texts, and asked subjects to practice this procedure with the display mask. The administrator then explained how to move from section to section of the experiment and allowed subjects to practice. Upon completion of this exercise, subjects prepared to begin the experiment.

Subjects moved to the first text, read their start time from the on screen timer, recorded their start time on the data form and immediately began reading. They read text 1 until completed, read their completion time from the timer, and recorded completion timeon the data form. Subjects then moved to the comprehension test, and answered four multiple choice (five choices) comprehension questions on that passage without referring back to the text.

When all subjects completed the questions, they recorded start time for text 2, read the treatment, recorded stop time, and answered comprehension questions in the same manner as before.

Data analysis

Each subject's personal information, reading times and comprehension scores were entered into a Microsoft Excel spreadsheet. The program organized raw scores into values for serif and sans serif texts, calculated reading rates, and evaluated comprehension scores. It also calculated each subject's reading rate difference between serif texts and sans serif texts, in words per minute. Reading time for the serif face was given a value relative to that of the sans serif, as sans serif is the industry standard for computer displays.

From this raw data table, scores for subjects who were unable to read the display, had computer-use problems, or scored poorly on comprehension questions were removed. The data was then entered into StatsView, a statistical software package, for further analysis.

Primary experimental effect

The primary experimental effect, the difference between subjects' serif reading rates and sans serif reading rates, was tested with a paired *t*-test with a critical value of $p<.05$.

Validity checks

The experimental design allowed for testing if type face caused all variations in subject performance. If this was the case, all the differences in reading rates for experimental treatments would be significantly different than those for the control treatment, and not significantly different than each other. Reading rate differences for the five treatments were examined with an analysis of variance. An unpaired *t*-test compared the reading rate differences of all the experimental treatments against that of the control.

Each subject's personal information (age, use of glasses or contact lenses, previous work with computers) was first compared with reading rates and reading rate differences, using correlation. An analysis of variance also checked for differences due to the class and time of day of experiment. Familiarity effects caused by order of reading of texts and order of reading of type faces were tested with *t*-tests.

Results

The experiment found no significant legibility difference between the serif and sans serif types. A paired t test comparing reading rates for serif and sans serif texts found t (42 df)=-1.951, p =.0577 (two-tailed). Statistical analysis of reading rates for the serif and sans serif texts failed to disprove the null hypothesis that "there is no significant difference between the reading speed of texts set in serif type faces and the reading speed of texts set in sans serif type face," at a probability level of p <.05.

Calculations of serif reading rates minus sans serif reading rates appear mostly negative, suggesting a trend of sans serif texts being read faster than serif. Indeed, the mean difference between the two type faces was -14.75 words per minute, in favor of the sans serif type face.

Threats to validity

The reading rates of serif and sans serif type faces are not significantly different, and validity checks suggest that the rate differences present are not necessarily due to differences between the type faces. The experiment was designed to allow for indirect measure of confounding variables in the experiment. Treatments 1 through 4 presented the subject with one serif and one sans serif text. These treatments varied only in the order of presentation of text and type face. Treatment 5, the control, presented two texts set in the same type face.

If the difference between serif and sans serif reading rates was caused exclusively by type face, the four experimental treatments would show a consistent difference between the serif and sans serif rates. The control treatment would show little difference between rates, as both texts were set in the same type face. Therefore, an unpaired *t*-test was used to look for a significant difference between the reading performance of subjects with experimental treatments and the reading performance of subjects with the control. The test found t (53 df)=-.412, p=.6818 (two-tail) showing that there is not significant difference between the experimental and control treatments. The lack of significant difference between experimental and control treatments suggests that type face is not the primary cause of reading rate differences in this experiment.

Analysis of reading rates of each treatment suggests that type face is not responsible for the majority of the experimental variance. Table 2 shows that in Treatments 1 and 2 serif texts were read much slower than sans serif texts, while in Treatments 3 and 4 serif texts were read at a similar rate or faster than sans serif texts. Analysis of variance results shown in Table 3 found a significant difference between treatments, a phenomenon which should not be present if variance was due to type face differences. Figure 8 showing treatment means and distributions suggests that the range of values in each treatment is so large that mean scores may lack descriptive value.

Table 2
Difference between serif and sans serif reading rates of treatments

Treatment	1st reading		2nd reading		*n*	Mean
	font, text		*font, text*			
1	serif	A	sans	B	11	-39.874
2	sans	A	serif	B	12	-25.033
3	serif	B	sans	A	10	16.18
4	sans	B	serif	A	10	- 5.597
5	sans	A	sans	B	13	- 6.446

Table 3
Analysis of variance between treatments

Tx	3	4	5	2	1
X rate	16.18	-5.59	-6.44	-25.03	-39.874

F(55 df)=2.736, *p*=.0387

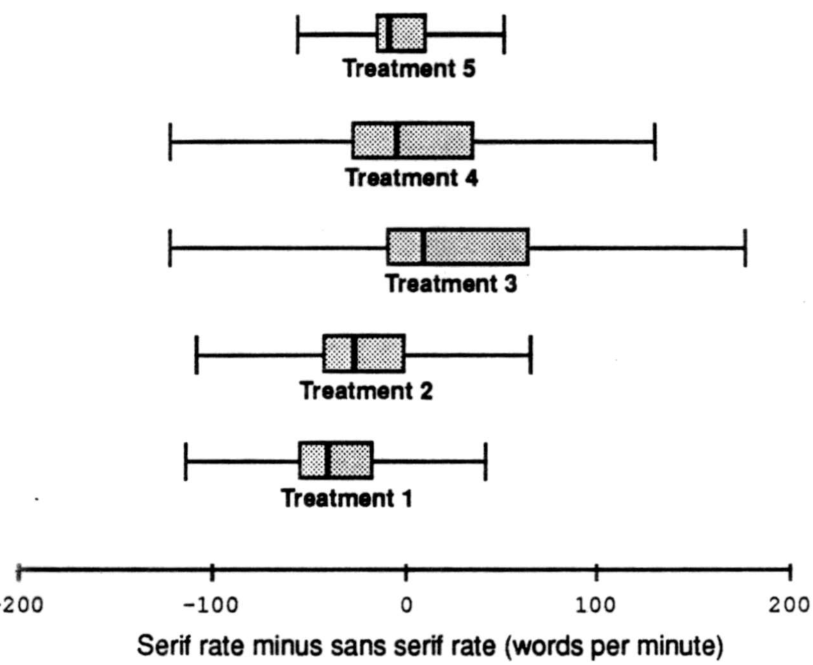

Figure 8
Range and distribution of scores of experimental treatments.

Subject factors

Age had no significant effects on reading rates for serif or sans serif treatments, or on the differences between the two rates. Subjects' use of prescription glasses or contact lenses had no significant effects on differences in serif and sans serif rates. Correlations and t tests found no significant effects of either of these variables on reading performance.

The number of hours subjects spent working on computers prior to the experiment likewise did not influence reading performance. Only 10 of the 56 subjects had worked on computers for an hour or more prior to the experiment. This number is too small to allow for proper consideration of the effect of visual fatigue on legibility of CRT displays.

Subjects were tested in four separate classes meeting on different days at different times, posing the potential that factors of time of day of testing, or makeup of class might affect the experiment's results. Analysis of variance tests of the reading rate differences of subjects in the different classes (shown in Table 4) found no significant differences between classes.

Table 4
Analysis of variance of class effects on differenc between serif and sans serif reading rates.

Class	2	4	3	1
Mean	-2.36	-14.12	-23.59	-24.66

F(42df)=.504, *p*=.6816(two-tail)

The mechanics of this experiment necessitated the comparison of reading rates from two separate texts similar in length and difficulty. Paired *t*-tests show no significant differences between average reading rates for text A and text B (*t* (53df)=-1.065, *p*=.2932 two-tail). Reading rates for subjects in Treatment 5, the control, which set both texts in the same type face, support this conclusion, as the reading rates for both texts were not significantly different, based on paired *t*-tests (*t* (53df)=-.399, *p*=.6935, two-tail).

Order of reading of passages may account for the variance between experimental treatments. Treatments 1 and 2 were significantly different than Treatments 3 and 4. Both 1 and 2 presented text A as the first reading and text B as the second. Treatments 3 and 4 presented the opposite arrangement. Analysis of the individual texts with a *t*-test showed that the reading rate for each **individual** text was not significantly different whether it was read first or second. While order of reading had no significant effects on performance of individual texts, order of reading of the **combination** of the two texts may have had an impact on reading performance.

Difference between the texts in content, vocabulary, or comprehension questions may have caused the experimental variation. Assume, for example, the comprehension questions following text A were more difficult than those following text B. Subjects reading text A first and finding difficult comprehension questions might then read text B differently because of the difficulty of the comprehension questions. Conversely, subjects first reading text B (with easier comprehension questions) would read text A similarly. If this were the case, Treatments 1 and 2 would differ from Treatments 3 and 4.

Discussion

While serif and sans serif text reading rates were not significantly different at $p<.05$ the experiment does show several trends. Sans serif texts were read an average of 14.72 words per minute faster than serif texts. This mean difference is significant if the probability level is raised to $p<.1$. Thus, with a less stringent confidence level, there is a significant difference between reading of serif and sans serif texts.

This observation may not be of value, however, because analysis of confounding variables suggests that type face was not necessarily the only variable affecting reading speed. Three considerations support the finding that reading rates of the two type faces are not significantly different.

1. The experimental treatments were not significantly different than the control.
2. The experimental treatments were significantly different than each other.
3. Order of reading appears to influence reader performance, explaining the differences between the experimental treatments.
4. Confidence levels of $p<.05$ have been used for most similar legibility experiments.

If type face differences cause any differences in legibility, this experimental design was too crude to detect these differences. Subjects timing their own reading surely introduced some error. Minor differences in scrolling through texts may also have altered scores. Subjects likely read the second text differently after completing the first and answering the comprehension questions. Subjects may also have compensated for legibility differences by adapting their vision (straining) more for less legible type faces. A prolonged (1-4 hour) reading rate experiment such as was done be Kruk and Muter (1984) and Mourant et al (1981) might rule out these factors from legibility.

The assignment of treatments to computer in the lab may also confound results. While treatments were randomly assigned to computers, the pattern of assignment may have caused an edge effect to contaminate the results. As shown in Figure 9, Treatments 3 and 4 were predominantly assigned to computers on the outside edges of the laboratory, while Treatments 1 and 2 were assigned to computers in the middle of the laboratory. Computers displaying Treatments 1 and 2 were therefore more likely to be positioned between two other computers, while Treatments 3 and 4 were more likely to be positioned with only 1 computer next to them. If glare or other distracting factors from neighboring computers decreased subject performance, then treatments 1 and 2 would be more subject to these effects.

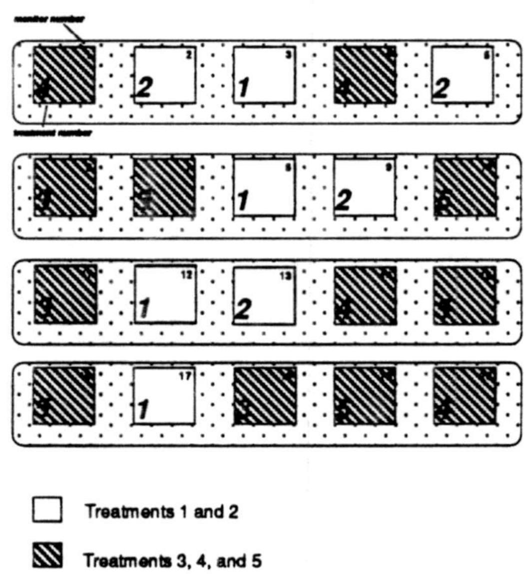

Figure 9
Edge effect caused by assignment of treatments to computers

Another noteworthy trend appears in this experiment. Subjects read the experimental texts at an average rate of 171.4 words per minute. This value is significantly slow for college readers with this type of text. If, as Gould and Grischkowski (1984) and Gould et al (1987) suggest, reading from CRT displays is 20-30% slower than from printed text, then we would expect subjects' reading rates of print to be 200-230 words per minute. Even this value is low for college readers.

Pretesting subjects reading rates of printed text prior to conducting this experiment would give insight into this trend. The use of relatively small type faces in the experiment is likely the cause of this slow reading rate. Texts were set in 9 point and 10 point type, at the smaller end of suggested type sizes for CRT displays. Repetition of the experiment with larger type sizes might yield a more precise measure of "natural" reading.

This study failed to conclude whether serif or sans serif type faces are better for use in computer displays. It does not, however, conclude that the two type faces are equally legible. Confounding variables in the experiment may have overwhelmed any effect (or lack of effect) that type face may have had. Further studies measuring subject's reading rate one at a time in a more controlled laboratory setting might better resolve this question.

The experimental design used for this study may not be appropriate for legibility research of this type. An experimental design such as Tinker's (1963) might a better measure reading rate and avoid the complications of order of reading effects. Further studies should attempt to recreate Tinker's reading rate instrument, the Minnesota Speed of Reading Test (out of print) which successfully measured reading performance in previous studies. Comprehension questions likely contaminated this study's results, and the Tinker test assures comprehension differently. Furthermore, use of this instrument might allow for more comparison of results to those of previous studies.

Studies combining these results with more sensitive types of experiments such as the reaction time experiments used previously in CRT legibility studies would further examine the issue. While speed of reading of continuous text is a very valid measure of legibility, it is not the only measure used. Indeed, the majority of CRT legibility studies use other methods. Comparing the results of this experiment to those of another method testing similar type faces might offer insight to the issue of appropriate type face.

Literature Cited

Aldus Corporation (1987). *Pagemaker Users Manual, Second edition.* Seattle: Aldus Corporation.

Atkinson, N.H., Crimley, L.M.,, & Willis, M.P. (1952). *A study of the requirements for letters, numbers, and markings to be used on trans-illuminated aircraft control panels.* (Report NAM EL-609), Philadelphia, PA: Aeronautic Medical Equipment Laboratory, Naval Air Experimental Station.

Baskette, F.K., Sissors, J.Z.,, & Brooks, B.S. (1986). *The art of editing.* New York: Macmillan Publishing Co.

Bentdy, M. (1921). Leading and legibility. *Psychological Monographs 30,* 48-61.

Brown, F.R. (1953). *A study of the requirements for letters, numbers and markings to be used on trans-illuminated aircraft panels. Part 4: Legibility of uniform stroke capital letters as determined by size and height-to-width ratio and as compared to Garamond Bold.* (Report TED No. NAMEL-609, Part 4). Philadelphia, PA: Aeronautic Medical Equipment Laboratory, Naval Air Experimental Station.

Brown, J.I., Bennett, J.M., & Hanna, G. (1981). *Nelson-Denny Reading Test. Forms E and F.* Chicago: Riverside Publishing.

Burt, C. (1953). *A psychological study of typography.* Cambridge: University Press.

Burtt, H.E., & Basch, C. (1923). Legibility of Bodini, Baskerville Roman, and Cheltenham type faces. *Journal of Applied Psychology 7,* 237-245.

Cakir, A., Hart, D.J., & Stewart, T. (1980). *Visual Display Terminals.* New York, John Wiley and Sons.

Carter, R.C. (1982). Search time with a color display: Analysis of distribution functions. *Human Factors 24,* 203-212.

Carter, R.C., & Cahil, M.C. (1979). Regression models of search time for color-coded information displays. *Human Factors 21,* 293-302.

Cheetham, D., & Grimbly, B. (1964). Design analysis: typeface. *Design 186,* 61-71.

Chen, H., & Tsoi, K. (1988). Factors affecting the readability of moving text on a computer display. *Human Factors 30,* 25-33.

Chen, H., Chan, K., & Tsoi, K. (1988). Reading self-paced moving text on a computer display. *Human Factors 30*, 285-291.

Christ, R.E. (1975). Review and analysis of color coding research for visual displays. *Human Factors 17*, 542-570.

Coleman, E.B., & Kim, L. (1961). Comparison of several styles of typography in English. *Journal of Applied Psychology 45*, 262-267.

Creed, A., Dennis, I., & Newstead, S. (1987). Proof-reading on VDUs. *Behavior and Information Technology. 6(1)*, 3-13.

Dowding, G. (1959). *Factors in the choice of type faces.* London: Wace and Co.

Duchnicky, R.L., & Kolers, P.A. (1983). Readability of text scrolled on visual display terminals as a function of window size. *Human Factors 25*, 683-692.

Duncan, T., & Konz, A. (1976). Legibility of LED and liquid crystal displays. *Proceedings of the S.I.D. 17*, 180-186.

English, E. (1944). A study of the legibility of four newspaper headline types. *Journalism Quarterly 21*, 217-229.

Felkner, D., Pickering, V.R., Charrow, V.M., Hollard, V.R., & Redish, J.C. (1981). *Guidelines for document designers*. Washington, D.C.: American Institutes for Research.

Fenton, E. (1989). At your service. *Macworld 4(10)*, 154-159.

Gould, J.D. (1968). Visual factors in the design of computer controlled CRT displays. *Human Factors 10*, 359-376.

Gould, J.D. & Grischkowski, N. (1984). Doing the same work with hard copy and Cathode-ray Tube (CRT) computer terminals. *Human Factors 26*, 323-337.

Gould, J.D., Alfaro, L., Barnes, V., Rinn, R., Grischkowsky, N., & Minuto, A. (1987). Reading is slower from CRT displays than from paper: Attempts to isolate a single-variable explanation. *Human Factors 29*, 269-299.

Granaas, M., McKay, T., Laham, R., Hurt, L., & Juola, J. (1984). Reading moving text on a CRT screen. *Human Factors 26*, 97-104.

Hackman, R.B., & Tinker, M.A. (1957). Effect of variations in color of print and background upon eye movements in reading. *American Journal of Optometry and Archives of the American Academy of Optometry 34*, 354-359.

Holmes, G. (1931). The relative legibility of black print and white print. *Journal of Applied Psychology 15*, 241-247.

Huddleston, H.F. (1974). A comparison of two 7x9 alphanumeric designs for TV display. *Applied Ergonomics 52*, 81-83.

Hvistendahl, J.K., & Kahl, M.R. (1975). Roman v. sans serif body type: Readability and reader preference. *ANPA News Research Bulletin 2*, 3-11.

IBM, (1987a). *IBM Personal System/2 Color Display 8513*. Rye Brook, N.Y.: IBM Information Systems Group.

IBM, (1987b). *IBM Personal System/2 Color Display 8503*. Rye Brook, N.Y.: IBM Information Systems Group.

IBM, (1987c). *IBM Personal System/2 Color Display 8512*. Rye Brook, N.Y.: IBM Information Systems Group.

IBM, (1987d). *IBM Personal System/2 Color Display 8514*. Rye Brook, N.Y.: IBM Information Systems Group.

International Paper. (1988). *The Pocket Pal*. New York: International Paper.

Jha, S., & Daftuar, C. (1981). Legibility of type faces. *Journal of Psychological Research 25*, 108-110.

Koler, P.A., Duchnicky, R.L., & Ferguson, D.C. (1981). Eye movement measurement of readability of CRT displays. *Human Factors 23*, 517-527.

Kruk, R.S., & Muter, P. (1984). Reading of continuous text on video screens. *Human Factors 26*, 339-345.

Krulee, G.K., & Novy, F. (1986). Word processing, & effects of variability in type fonts. *Perceptual and Motor Skills 62*, 999-1010.

Luckeish, M., & Moss, F.K. (1935). Visibility: Its measurement and significance in seeing. *Journal of the Franklin Institute 220*, 431-466.

Luckeish, M., & Moss, F.K. (1937) The visibility of various type faces. *Journal of the Franklin Institute 223*, 77-82.

Maddox, M.E., Burnette, J.T., & Gutmann, J.C. (1977). Font comparison for 5x7 dot matrix characters. *Human Factors 19,* 89-93.

Matula, R.A. (1981). Effects of visual display units on the eyes: A bibliography (1972-1980) *Human Factors 23*, 581-586.

McVey, G.F. (1985). Legibility in film based and television display systems. *Technical Communication 32,* 21-27.

Microsoft Corporation (1987). *Microsoft Windows Users Guide, Version 2.0.* Seattle: Microsoft Corporation.

Mourant, R., Ladshmanan, R.,, & Chautadisai, R. (1981). Visual fatigue and cathode ray tube display terminals. *Human Factors 23* , 529-540.

Murrell, K. (1965). *Ergonomics*. London: Chapman and Hall.

Muter, P. Latremouille, S., & Treurniet, W. (1982). Extended reading of continuous text on television screens. *Human Factors 24* , 501-508.

Nilson, L. Ronnberg, J., & Ohlsson, K. (1983). Ratings of color combinations of text and background presented by means of a videotext system. *Umea Psychological Reports 166*, 1-24.

Paterson, D.G., & Tinker, M.A. (1929). Studies of typographical factors influencing speed of reading: II. Size of type. *Journal of Applied Psychology 13*, 120-130.

Paterson, D.G., & Tinker, M.A. (1931). Studies of typographical factors influencing speed of reading: VI. Black type versus white type. *Journal of Applied Psychology 15*, 241-247.

Paterson, D.G., & Tinker, M.A. (1932). Studies of typographical factors influencing speed of reading: IV. Space between lines or leading. *Journal of Applied Psychology 16*, 388-397.

Paterson, D.G., & Tinker, M.A. (1940a). *How to Make Type Readable*. New York: Harper and Brothers.

Paterson, D.G., & Tinker, M.A. (1940b). Influence of line width on eye movements. *Journal of Experimental Psychology 27*, 572-577.

Paterson, D.G., & Tinker, M.A. (1942). Influence of line width on eye movements for six point type. *Journal of Educational Psychology 33*, 552-555.

Pittman, B. (1976). Style of type and the reading comprehension of learning disabled and normal pupils. *Journal of Research and Development in Education 9*, 114-115.

Poulton, E.C. (1965). Letter differentiation and rate of comprehension in reading. *Journal of Applied Psychology 49*, 358-362.

Pyke, R.K. (1926). *Report on the legibility of print.* London: H.M. Stationery Office.

Rehe, R.F. (1974). *Typography: How to make it most legible.* Chapel Hill, N.C.: Source Publications.

Riley, T.M., & Barbato, G.D. (1978). Dot matrix alphanumerics viewed under discreet element breakdown. *Human Factors 20*, 473-479.

Robinson, D.O., Abbamonte, M., & Evans, S.H. (1971). Why serifs are important: The perception of small print. *Visible Language 4*, 353-359.

Roethlein, B.E. (1912). The relative legibility of different faces of printing type. *The American Journal of Psychology 23*, 1-36.

Roland, G.E., & Cornog, D.Y. (1958). *Selected alphanumeric characteristics for closed circuit television displays.* (Technical Report No. 21). Philadelphia: Courtney and Company.

Rubens, P., & Krull, R., (1985). Application of research on document design to online displays. *Technical Communication 32*, 29-34.

Sanocki, T. 1987. Visual knowledge underlying letter perception: Font-specific, schematic tuning. *Journal of Experimental Psychology: Human Perception and Performance 13*, 267-278.

Shurtleff, D.A. (1974). Legibility research. *Proceeding of the S.I.D. 15(2)*, 41-52.

Shurtleff, D.A., & Owen, D. (1966). *Studies of display symbol legibility: VI. A comparison of the legibility of televised Leroy and Courtney Symbols.* (Technical Report ESD-65-136). Bedford, MA: Hascom AFB.

Shurtleff, D.A., Marsetta, M., & Showman, D. (1966) *Studies of display symbol legibility:IX. The effects of resolution, visual size an viewing angle on the legibility of televised Leroy alphanumeric symbols.* (Technical Report ESD-65-411). Bedford, MA: Hascom AFB.

Shurtleff, D.A. (1980). *How to make displays legible.* La Mirado, CA: Human Interface Design.

Smith, S.L. (1979). Letter size and legibility. *Human Factors 21*, 661-670.

Tinker, M.A. (1931). Apparatus for recording eye movements. *American Journal of Psychology, 43*, 115-118.

Tinker, M.A. (1936). Reliability and validity of eye-movement measures of reading, *Journal of Experimental Psychology, 19*, 732-746.

Tinker, M.A. (1944). Criteria for determining the readability of type faces. *Journal of Educational Psychology, 35*, 385-396.

Tinker, M.A. (1945). Reliability of blinking frequency employed as a measure of readability. *Journal of Experimental Psychology, 35,* 418-424.

Tinker, M.A. (1947). Time relations for eye movement measures in reading. *Journal of Educational Psychology, 38*, 1-10.

Tinker, M.A. (1948). Readability of book print and newsprint in terms of blink rate. *Journal of Educational Psychology, 39*, 35-39.

Tinker, M.A. (1950). Reliability and validity of involuntary blinking as a measure of ease of seeing. *Journal of Educational Psychology 41*, 417-427.

Tinker, M.A. (1963). *The legibility of print.* Ames, Iowa: University of Iowa Press.

Tinker, M.A., & Paterson, D.G. (1928). Influence of type form on speed of reading. *Journal of Applied Psychology 12*, 359-368.

Tinker, M.A., & Paterson, D.G. (1931). Studies of typographical factors influencing speed of reading: VII. Variations in color of print and background. *Journal of Applied Psychology 15*, 471-479.

Tinker, M.A., & Paterson, D.G. (1942). Reader preferences and typography. *Journal of Applied Psychology 26*, 38-40.

Tinker, M.A., & Paterson, D.G. (1943). Differences among newspaper body types in readability. *Journalism Quarterly 20*, 152-155.

Tullis, T.S. (1981). An evaluation of alphanumeric, graphic, and color information displays. *Human Factors 23*, 541-550.

Tullis, T.S. (1983). The formatting of alphanumeric displays: A review and analysis. *Human Factors 25,* 657-682.

Vartebedian, A.G. (1971a). Legibility of symbols on CRT displays. *Ergonomics 2*, 130-132.

Vartebedian, A.G. (1971b). The effects of letter size, case, and generation method on CRT display search time. *Human Factors 13*, 363-368.

Webster, J., & Tinker, M.A. (1935) The influence of type face on the legibility of print. *Journal of Applied Psychology 19*, 43-52.

Wilkins, A. (1986). Intermittent illumination from visual display units and fluorescent lighting affects movements of the eyes across text. *Human Factors 28*, 75-81.

Wright, P., & Lickorish, A. (1983). Proofreading texts on screen and paper. *Behavior and Information Technology 2*, 227-235.

Appendices

Appendix A:
Nelson Denny Reading Test Statistics

Validity and Reliability

Texts used for this experiment were extracted from the Nelson-Denny Reading Test, Forms E and F . Reading texts ranked at a 8th to 9th grade difficulty level by Flesch scales, and a 9th to 10th grade difficulty level by Corrected Dale-Chall Grade Level tests. (Brown et al, 1981).

Comprehension Questions

Comprehension questions for passages were developed by pretesting 25 possible questions for each passage. The four most valid and reliable questions for each passage, were chosen from the pool of 25 possible questions through pre-testing prior to publication of the Nelson-Denny Reading Test (Brown et al, 1981).

Appendix B:

Passages used in experiment

With increasing prosperity, West European youth is having a fling that is creating distinctive consumer and cultural patterns.

The result has been the increasing emergence in Europe of the pehnomenon well known in America as the "youth market." This is a market in which enterprising businesses cater to the demands of teenagers and older youths in all their rock mania and pop-art forms.

In the United States, the market is wide-ranging and well established, almost an industry, which with this country's emphasis on "youthfulness," even extends beyond teen-ager groups.

In Western Europe, the youth market may appropriately be said to be in its infancy. In some countries such as Britain, West Germany and France, it is more advanced than in others. Some manifestations of the market, chiefly sociological, have been recorded, but it is only just beginning to be the subject of organized consumer research and promotion.

Characteristics of the evolving European youth market indicate dissimilarities to the American youth market.

The similarities:

The market's basis is essentially the same--more spending power and freedom to use it in the hands of teen-agers and older youth. Young consumers also make up an increasingly high proportion of the population.

As in the United States, youthful tastes in Europe extend over a similar range of products--records and record players, transistor radios, leather jackets and "wayout," extravagantly styled clothing, cosmetics, and soft drinks. Generally it now is difficult to tell in which direction trans-Atlantic teen-age influences are flowing.

Also, a pattern of conformity dominates European youth as in this country, though in Britain, the object is to wear clothes that make the wearer stand out, but also make him "in," such as tight trousers and precisely tailored jackets.

Youth worship and emulate "idols" in the entertainment field, especially the "pop" singers and other performers. There is also the same exuberance and unpredictability in sudden fad switches. In Paris, buyers of stores catering to the youth market carefully watch what dress is being worn by a popular television teen-age singer to be ready for a sudden demand for copies. In Stockholm other followers of teen-age fads call the youth market "attractive but irrational."

As in the United States where "teen" and "teener" have become merchandising terms, Europeans also have adopted similar terminology. In Flemish and Dutch it is "tiener" for teen-agers. In West Germany the key word in advertising addressed to teen-agers is "freizeit" meaning holidays or time off.

The most obvious differences between the youth market in Europe and that in the United States is in size. In terms of volume and variety of sales, the market in Europe is only a shadow of its American counterpart, but it is a growing shadow.

In West Germany, for example, teenagers now are recognized as accounting for 10 percent, or $3 billion, of retail sales a year.

Actually, the scope and nature of the youth market varies considerably from country to country, being large and lively in some and only beginning to show itself in others.

But there are also these important dissimilarities generally with the American youth market:

In the European youth market, unlike that of the United States, it is the working youth who provides the bulk of purchasing power.

On the average, the school-finishing age still tends to be 14 years. This is the maximum age to which compulsory education extends, and with Europe's industrial manpower shortage, thousands of teen-age youths may soon attain incomes equal in many cases to that of their fathers.

Although, because of general prosperity, European youths are beginning to continue school studies beyond the compulsory maximum age, they do not receive anything like the pocket money or "allowances" of American teen-agers. The European average is about $5 to $10 a month.

Working youth, consequently, are the big spenders in the European youth market, but they also have less leisure than those staying on at school, who in turn have less buying power.

1. The main purpose of the passage is to :
   A) discuss new trends in the consumer patters of youth
   B) describe different types of clothing worn by youth
   C) analyze several advertising techniques used in gaining the youth market
   D) compare the youth in the United States with the youth of Europe
   E) show how economic conditions in Europe affect sales of goods to American youth

2. The passage states or implies that the youth market is more established in:
   A) France than it is in West Germany
   B) Italy than it is in France
   C) Belgium than it is in Sweden
   D) England than it is in our country
   E) our country than it is in Europe

3. According to the passage, the youth market is one that :
   A) is exceptionally easy to predict
   B) is responsible for a larger percentage of consumption than was true a few years ago
   C) has long been under the scrutiny of consumer investigators
   D) is likely to taper off in the next decade
   E) has never been clearly distinguished from the adult market

4. The American and European youth markets are alike in regard to :
   A) appreciation of the same individual entertainers
   B) amount of money per (teen-ager) capita spent on clothing
   C) occurrence of frequent changes in buying habits
   D) amount of parental allowances for clothing
   E) purchase of luxuries

Educators are seriously concerned about the high rate of dropouts among the doctor of philosophy candidates and the consequent loss of talent to a nation in need of PhDs. Some have placed the dropout loss as high as 50 percent. The extent of the loss was, however, largely a matter of expert guessing.

Last week a well-rounded study was published. It was based on 22,000 questionnaires sent to former graduate students and seemed to show many past fears to be groundless.

The dropout rate was found to be 31 percent, and in most cases the dropouts, while not completing the PhD requirements, went on to productive work.

They are not only doing well financially, but, according to the report, are not far below the income levels of those who went on to complete their doctorates.

The study, called *Attrition of Graduate Students at the PhD Level in the Traditional Arts and Sciences,* was made at Michigan State University under a $60,000 grant from the United States Office of Education. It was conducted by Dr. Allan Tucker, former assistant dean of the university and now chief academic officer of the Board of Regents of the State University System of Florida.

Discussing the study last week, Dr. Tucker said the project was initiated, "because of the concerns frequently expressed by graduate faculties and administrators that some of the individuals who dropped out of PhD programs were capable of completing the requirements for the degree.

"Attrition at the PhD level is also thought to be a waste of precious faculty time and a drain on university resources already being used to capacity. Some people expressed the opinion that the shortage of highly trained specialists and college teachers could be reduced by persuading the dropouts to return to graduate shcools to complete the PhD program."

"The results of our research," Dr. Tucker concluded, "did not support these opinions."

The study found that:

(1) lack of motivation was the principal reason for dropping out.

(2) most dropouts went as far in their doctoral programs as was consistent with their levels of ability or their specialities.

(3) most dropouts are now engaged in work consistent with their education and motivation.

(4) the dropout rate was highest in the humanities (50%) and lowest in the natural sciences (29%)--and is higher in lower-quality graduate schools.

Nearly 75 percent of the dopoouts said there was no academic reason for their decision, but those who mentioned academic reasons cited failure to pass qualifying examinations, uncompleted research and failure to pass language exams.

"Among the single most important personal reasons identified by dropouts for non completion of their PhD program," the study found, "lack of finances was marked by 19 percent."

As an indication of how well the dropouts were doing, a chart showed that 2 percent whose studies were in the humanities were receiving $20,000 and more annually, while none of the PhDs with that backgound reached this figure. The PhDs shown in the $7,500 to $15,000 bracket had 78 percent at that level against 50 percent for the dropouts. This may also be an indication of the fact that top salaries in the academic fields, where PhDs tend to rise to the highest salaries, are still lagging behind other fields.

In the social sciences 5 percent of the PhDs reached the $20,000 plus figure as against 3 percent of the dropouts, but in the physical sciences they were neck-and-neck with 5 percent each.

Academic institutions employed 90 percent of the humanities PhDs as against 57 percent of the humanities dropouts. Business and industry employed 47 percent of the physical science PhDs and 38 percent of the physical science dropouts. Government agencies took 16 percent of the social science PhDs and 32 percent of the social science dropout.

As to the possibility of getting dropouts back on campus, the outlook was glum.

"The main conditions which would have to prevail for at least 25 percent of the dropouts who might consider returning to graduate school would be to guarantee that they would retain their present level of income and in some cases their present job.

1. The author states that many educators feel that:
   A) steps should be taken to get the dropouts back to school particularly in certain disciplines
   B) since the dropout does just about as well financially as the PhD holder, there is no justifiable reason for the dropout to return to his studies
   C) the high dropout rate is largely attributable to the lack of stimulation on the part of faculty members
   D) the dropout should return to a lower-quality school to continue his studies
   E) the PhD holder is generally a better-adjusted person than the dropout

2. Research has shown that:
   A) dropouts are substantially below PhDs in financial attainment
   B) the incentive factor is a minor one in regard to pursuing PhD studies
   C) the PhD candidate is likely to change his field of specialization if he drops out
   D) about one third of those who start PhD work do not complete the work to earn the degree
   E) there are comparatively few dropouts in the PhD humanities disciplines

3. Meeting foreign language requirements for the PhD :
   A) is the most frequent reason for dropping out
   B) is more difficult for the science candidate than for the humanities candidate
   C) is considered part of the so-called "qualification" exam
   D) is an essential part of many PhD programs
   E) does not vary in difficulty among universities

4. Dr. Tucker felt that :
   A) a primary purpose of his research project was to arrive at a more efficient method for dropping incapable PhD applicants
   B) a serious aspect of the dropout situation was the deplorable waste of productive talent
   C) one happy feature about the dropout situation was that the dropouts went into college teaching rather than into research
   D) his project should be free of outside interference and so he rejected outside financial asistance for the project
   E) it was important to determine how well PhD dropouts did in comparison to those who completed the program