Running Head: LEGIBILITY AND COMPREHENSION OF ONSCREEN TYPE

Comparing the Legibility and Comprehension of Type Size, Font Selection and Rendering Technology of Onscreen Type

by Scott Bondurant Chandler

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University,
Blacksburg, Virginia,
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Curriculum and Instruction

APPROVED:

Glen Holmes, PhD, Co-chair

Peter Doolittle, PhD, Co-chair

Tom Hergert, PhD, Member

Mark Sanders, PhD, Member

Ken Potter, PhD, Member

Keywords: type, typography, legibility, speed of reading, reading speed, comprehension, education, effectiveness, type size, font selection, serif, sans-serif, type rendering technology, anti-alias, perceptual encoding, orthochromatic, Helvetica, Palatino, computer, CRT, monitor, experimental, repeated measures.

Comparing the Legibility and Comprehension of Type Size, Font Selection and Rendering Technology of Onscreen Type

by Scott Bondurant Chandler

Abstract

This experimental study investigated the relationship between the independent measures of font selection, type size, and type rendering technology and the dependent measures of legibility, as measured by the Chapman-Cook speed of reading test, as well as comprehension, as measured by a series of questions from the verbal comprehension section of the Graduate Record Exam.

An electronic instrument presented test items in 12 different typographic styles. The study tested 117 college students at a university in southwestern Virginia. Each participant encountered anti-alias type rendering style and the orthochromatic type rendering style while participants were randomly assigned to either Helvetica or Palatino (font selection) and 8, 10 or 12 point type size.

Results indicated that the 12 point type size was read more quickly than either 8 point type or 10 point type. There was also an interaction between font selection and type rendering technology for speed of reading: Helvetica without an anti-alias was read more quickly than Helvetica with an anti-alias and more quickly than Palatino without an anti-alias. These findings contradict an earlier, similar study.

There were no significant results with regard to comprehension.

Keywords: type, typography, legibility, speed of reading, reading speed, comprehension, education, effectiveness, type size, font selection, serif, sans-serif, type rendering technology, anti-alias, perceptual encoding, orthochromatic, Helvetica, Palatino, computer, CRT, monitor, experimental, repeated measures.

Acknowledgements

"I am the vine; you are the branches. If a man remains in me and I in him, he will bear much fruit; apart from me you can do nothing." John 15:5

Thanks to you, Kelly. I appreciated your support and help every step of the way. Beyond what was required, you seemed to innately know what to say and how to encourage. I could not have done it without you.

With sincere appreciation to my supportive committee members (Glen Holmes, Peter Doolittle, Greg Sherman, Mark Sanders, Tom Hergert, Ken Potter, Larnie Cross) and to my teachers, all of them: Bijward, Price, Kelly, Lemon, Svec, Arthur, Barnett, Nuff, Wimberly, Wardell and Higgins to name only a few.

Kelly Wolff, Jason Hutchens, and Matt Gardner provided substantial editorial and technical support. Kathy Pinkerton and Bob Fields provided crucial access to their students, most of whom volunteered to partipate. Susan Ariew was more than just a fellow researcher: she was an advocate.

And thanks to supportive friends and family, you know who you are.

Table of Contents

List of Tables

List of Figures

Introduction

Problem Presentation

There is a rich history of research regarding typography in print. One needs look no further than the seminal works of Miles Tinker or thumb through the pages of the *Journal of Applied Psychology* around the 1940s to find hundreds of studies on very specific issues relating to the use of type.

More recently, much attention has been paid to the educational use of type as well as structuring and presentation of text in print format. Common threads have developed and most researchers agree on how the effectiveness of text information can be maximized for training and education (Glynn & Britton, 1984; Kostelnick, 1990; Kramer & Bernhardt, 1996; Streit, Davis, & Ladner, 1986).

Although many sources discuss the layout and selection of typefaces in print, very few studies have experimentally addressed this issue on the computer screen (Waite, 1995) and much of what research does exist is dated (Jones, 1994). With the educational trend towards Internet instruction and interactive multimedia, there seems to be a gap in the visual literacy research.

It is possible that certain styles, sizes and presentations of type could affect learning. If such results were found, the potential exists to improve the screen designs of interactive multimedia and web-based instruction. Practical suggestions for the use of on-screen type could enhance and improve instructional multimedia as type design is often handled by programming teams and not by graphic designers (Mukherjee & Edmonds, 1993).

Traditional wisdom has suggested that serif typefaces are easier to read than sans-serif faces. Typographer Ruari McLean's first rule of typographic legibility is that "sans-serif is intrinsically less legible than serifed type" (McLean, 1997). Of course, it has yet to be proved whether this rule holds true in onscreen applications.

Historically, some letters were distinguishable only by their serifs. The German blackletter fonts from the 1400s are a good example of this phenomenon. Hoener, Salad and Kay (1997) report that research on typeface selection has been inconclusive. However, this study used typeface in a different way. It is possible that although serif and sans-serif typefaces are equally legible in the relatively high-resolution of printed instruction, the low-resolution of the computer screen and mode of onscreen presentation may indicate a preference.

The typographic community believes that small type is considered more elegant and leaves room for negative space, which is thought to improve the overall legibility of the page. In fact, traditional lead type was cast and set as small as a nonpareil, now commonly known as six point type (Bringhurst, 1999). Perhaps the lower resolution (image density) of interactive multimedia makes small type too difficult to read.

This study addresses type rendering technology. There is a trend towards anti-alias rendering (grayscale or tonal) instead of orthochromatic rendering (pure black and pure white) of type on computer screens (Gardner, 2000; Fields, 2000; Gowen, 2000). Although Wiessenmiler (1999) implicitly addressed rendering technology in"A study of the readability of on-screen text," this area has received little attention in the research.

Overview

A wide range of literature needs to be addressed before undertaking this study. Much of the literature deals with traditional typography, which was based entirely on the printed word. In addressing the specific variables employed in this study, the immense body of knowledge with respect to type selection and type size in the realm of print will be considered. This understanding helps to form reasonable theories as starting points for researching onscreen type.

Much of the literature on non-print typography addresses early television-presented instruction. Later research addresses the screen design of text-only (ASCII) computer systems and low-resolution displays.

There is a small but relevant body of work dealing with type onscreen. While most of this work deals specifically with type on the world wide web, these articles directly impact the goals and structure of this study.

Definitions

For the purpose of this study, a number of technical terms need to be precisely defined and understood:

Font selection: the choice of typeface for use as body copy. Type selection is broken down into several broad categories including serif and sans-serif. Almost all serif of these typefaces include small accents at the edge of the letterforms (see Figure 1); these additions are thought to increase legibility in print. Sans-serif faces include contemporary faces in which the letterforms are more consistent in stroke and are not highlighted by

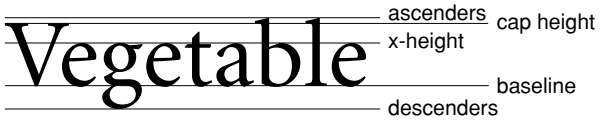
ornaments (see Figure 1). Historically the term typeface was used to refer to the look of a particular typographic design while the term font referred to one specific size of typeface expressed in metal (McLean, 1997). With the advent of computerized typesetting, the terms “typeface” and “font” are used interchangeably. In this study, both terms refer to a font design and not to a particular size.

Figure 1
Serif vs. sans-serif styles

Serifs are small accents at the ends of letterforms in historical typefaces. Contemporary sans-serif faces lack these “feet.”

Font size: the size in points (1/72 inch) from the top of the tallest letter in a typeface to the bottom of the descenders of the lowest lowercase letter with no additional spacing added (see Figure 2). Even at the same point size, different typefaces may be very different in size. Script faces, for example, have larger than normal descenders (lowercase letters dip below the baseline) making the actual size of the type smaller.

Figure 2
Typeface terminology regarding fount height



36 point Minion reveals how a typeface’s size is determined not by the height of most letters but of the most extreme letters.

Rendering technology: Different computer operating systems, software applications and font technologies use distinct methods for calculating the shape of letters. Therefore, there are many visibly different ways for text to appear onscreen. Historically most type on computer screens was rasterized—that is, rendered onscreen using square pixels on an orthogonal grid—using only pure black and pure white (orthochromatic mode); some more sophisticated systems now use subtle changes in value (grayscale) in an effort to improve the appearance of type. The use of grayscale in type rendering is known as anti-alias.

Legibility: refers to the relative ease with which individual letterforms, words and paragraphs may be read. Speed of reading has historically been used as one method to ascertain the legibility of different size and styles of type.

Readability: describes the complexity of the words that make up the message being read. Readability was once used interchangeably with legibility but has now taken on a different and broader meaning.

Comprehension: refers to the degree to which students are able to complete a cognitively difficult set of questions using recall and reasoning.

Review of Literature

History of Early On-screen Research

Television-Presented Graphics

During the 1980s, there were several television-based systems that provided a platform for computer based learning. Several of these showed promise and are discussed at length in the research: videotext and teletext were two of the most common technologies (Carey, 1984; Rubin, 1984). However, other "convergent" technologies replaced these broadcast methods. By the early 1990s, CD-ROM had become the dominant technology. Today it is common to use CD-ROM, Internet technologies such as the world wide web as distribution media for instruction employing text, graphics, animation, video and sound.

Computer-Presented Graphics:

Isaacs (1987) makes a strong case for the study of typography in multimedia applications in "Text Screen Design for Computer-Assisted Learning." The author suggests that although many macro areas of multimedia have been researched, many of the "small scale" issues such as text design have been overlooked.

Nonetheless, Issacs (1987) offers only a cursory overview of type design displayed on a computer monitor. The author does make obvious pronouncements. For instance, Issacs suggests that mixed-case text is easier to read than upper-case text and that color is a factor in the readability of type. Issacs' (1987) research experimentally determined that some colors are better than others, that reversed video may be difficult to read, that

blinking and flashing should be used sparingly, and that type style, size and line length all affect the legibility of the message.

It is important to realize that Isaacs' article was written in 1987. Options for the design of multimedia text were limited by current standards. In fact, the author defines an "ordinary colour display" as 320x200 pixels supporting 16 colors. Even the most inexpensive computer today dramatically exceeds those specifications.

Font Selection

Font Selection in Print

According to typographer Sean Cavanaugh "when setting body text… serif typefaces are naturally better than sans serif typefaces" (Cavanaugh, 1995, p. 105). Likewise, he reported that headlines and forms are appropriate uses for sans-serif faces. While few other typographers make such sweeping pronouncements, most typographers have a demonstrable bias towards the classical lines of the serif faces made popular by the Renaissance (Bringhurst, 1999; Cavanaugh, 1995). For example, the vast majority of books are printed in serif faces.

In his classic "Elements of Typographic Style," Robert Bringhurst (Bringhurst, 1999) urges extreme caution in selecting typefaces. For instance, he made a distinction between fonts designed for letterpress (metal type pressing into paper like Gutenberg) and offset press (ink transferred onto paper via a rubber blanket). Since laser printers operate at resolutions near 300 dpi, Bringhurst cautions against using delicate or modern stroke fonts at this resolution. Palatino (a serif face) and Optima (sans-serif face with strokes that vary in size with some parts of each glyph thicker and others thinner) are

explicitly named as troublesome. A computer screen has less than 1/17th the resolution (in area) of the laser printer Bringhurst shames.

Paterson and Tinker (1932) studied the relative speed of reading of the seven most frequently used typefaces. These seven faces included Scotch Roman, Garamond, Antique, Bodoni, Old Style, Caslon, Cheltenham. Of these seven faces, variations of Garamond, Bodoni, and Caslon are used today for design purposes while only Garamond is commonly used for body copy. The literature suggests that there is very little difference in legibility. These seven serif faces vary in speed of reading by a maximum of 2.8% (Donald Gildersleeve Paterson & Tinker, 1932). The study also included three display faces: Kabel Light, arguably a sans-serif face, read 2.6% slower than Garamond, a serif face; American Typewriter and Cloister Black read 5.1% and 14.0% slower than Garamond respectively (Donald Gildersleeve Paterson & Tinker, 1932; Tinker, 1963). Pyke (1926) suggests that there are few variations in serif faces and therefore only radical changes in the design of a typeface will result in appreciable differences in legibility. Other research confirms this premise when letters are formed into words and phrases but not when viewed separately (Rothlein, 1912).

Expert Opinion on Onscreen Font Selection

Issacs (1987) suggests that the design and size of type will affect legibility. The parameters of that research keep those observations from being useful with today's larger screens. Since Issac's research, there have been improvements in monitor technology. Today's screens offer more pixels, higher pixel resolution and improved color depth.

Additionally, all popular operating systems now sport a graphical user interface which provides a radically different user experience than what was standard in 1987.

Mason observes that serif typefaces and modern typefaces ("modern" refers to typefaces with variable stroke widths such as Bodoni and Optima) may be difficult to read on-screen (Mason, 1997; McLean, 1997; Williams, 1994). At small sizes these styles may lack the resolution to properly distinguish letterforms (Mason, 1997).

Research primarily focused on text density describes some of the variables considered here. For instance, Ipek (1995) states that "growing evidence suggests that many design principles are unique to the computer." Geske (1996) predicted that serif typefaces onscreen "may present unique problems" in that the serif fonts may not have adequate resolution to render the subtle character shapes and may darken when rendered. In general, Geske predicts that serif typefaces will not perform well. Other scholars concur with his premise (Bradshaw, 2000; Sutherland, 2000). Further, Bigelow states

> When printed the serifs on typefaces are only a tiny percentage of the typeface design. But on-screen, in order to display the serifs using the limited number of available pixels, they take up a much bigger proportion of the information than they do in print. Serifs should be small things—but on screen they become big… noise or distracting chunks of interference. (Bigelow in Petzgold, 1992)

Onscreen Font Selection Research

Nonetheless, Geske's (2000) recent study on the "Readability of body text in computer mediated communication" found that there were no statistically significant

differences in reading serif typefaces onscreen when compared to sans-serif. Geske used Palatino as his serif font and Helvetica as his sans-serif font. Although non-significant differences should be viewed with great caution, there is so little research on this topic that these data might warrant review. In 14 point size, serif faces were read faster (77.4 seconds compared to 81.6 seconds, $t > .05$); in 12 point size, reading times were similar (74.0 seconds for serif compared to 74.8 seconds for sans-serif, $t > .05$); and for 10 point size sans-serif was read somewhat more quickly (81.9 seconds versus 83.9 seconds, $t > .05$).

Weisenmiller (1999) compared four typefaces (Georgia, Times, Arial, and Verdana) in "A study of the readability of on-screen text." In addition to looking at comparing serif typefaces to sans-serif typefaces, Weisenmiller also considered one font of each type designed for use onscreen and one font of each type designed for use in print. However, it is likely that at least one of the print-based fonts was "hinted" (optimized with special programming for onscreen use). Weisenmiller found no differences in speed of reading or comprehension between any of the four typefaces.

Grant's (2000) recent study comparing serif vs. sans-serif type in testing environments presented via the world wide web found a significant difference based on font selection. This experiment indicated that serif type was read 67.9 words faster per minute than sans-serif type. This represents a 25% improvement over sans-serif type. Windows based users viewed Times New Roman (serif, TrueType) and Ariel (sans-serif, TrueType) while Macintosh based users viewed Times (serif, Postscript or TrueType) and Helvetica (sans-serif, Postscript or TrueType).

Essentially, Grant prepared an electronic version of the verbal comprehension component of the Graduate Record Exam using web-based technologies. Students were instructed to access the appropriate web page from their own computer in their usual computing environment. They were then asked to provide demographic information and begin the GRE test. The amount of time to read each passage was timed although the time needed to answer each question was not recorded. At the end of the test, the amount of time needed to read each passage and the answer for each passage was stored on the web server. Because of the random assignment of participants to different font styles a comparison of means could be conducted to compare serif and sans-serif faces.

Grant's study does not explain how or if differences between operating systems, font technologies, browser type, monitor settings or rendering style were controlled.

Common Practice in Onscreen Font Selection

Unlike in print, the fonts available on both the machine of the content creator and the machine of the content viewer often influence font selection onscreen. While some technologies—such as Adobe Acrobat, Adobe Photoshop, MacroMedia Director—preserve font style when transferred to a client machine, other technologies such as the world wide web and Microsoft Word do not.

Specific font selection onscreen is largely uncontrolled. In word processors and on the world wide web, the most used type specifications are likely to be the application defaults. Gardner (2000) suggests that 45% of all onscreen type is viewed in one of the variants of Times because it acts as the pre-programmed default in many applications.

A review of web sites (Gardner, 2000) confirms that web fonts are highly limited. Many web sites that have been around since the early days of the Internet, rely heavily on Times (Macintosh), Times New Roman (Macintosh and Windows), Helvetica (Macintosh) and Arial (Macintosh and Windows). Most font selection on the web today appears to be limited to a short list of TrueType fonts including Times New Roman, Arial, Tahoma, and Verdana (Gardner, 2000). Although sans-serif faces are more common, serif and sans-serif fonts are used in a variety of sizes and applications (see Table 1). It is also evident that the choice of browser plays a major role in the size at which any particular text element is drawn (Vosseller, 2000).

Table 1
Font Selection and Sizes in Web Sites

	navigation	subheads	body copy
yahoo.com (IE5)	serif, 12	sans, 16	serif, 16
yahoo.com (net47)	serif, 9	sans,12&13	serif, 10
altavista.com (IE5)	sans, 12	sans, 16	sans, 12
altavista.com (net47)	sans, 12	sans, 16	sans, 12
www.dell.com (IE5)	sans, 9	sans, 13	sans, 9&13
www.dell.com (net47)	sans, 9	sans, 9	sans, 9&10
macnn.com (IE5)	sans, 11	serif, 16	serif, 16
macnn.com (net47)	sans, 9	serif, 11	serif, 10

Font selection and size vary among web sites. In fact, meaningful differences exist between Microsoft Internet Explorer 5.0 (IE5) and Netscape Explorer 4.7 (net47) web browsers.

There also appears to be a trend that most textual navigation bars now use type rendered on the client machine instead of type pre-rendered as graphics by the web designer (Peck, 2000). While this compromises the exact appearance of the type, it is likely to increase the speed at which the page loads (Peck, 2000).

Type Size

Introduction to Type Size

The current system of point size is an anachronism. To fully appreciate the system of measurements, terminology and use of typography one must understand the history of type dating back to individual letterforms made from hot metal.

Typeface sizes are customarily described in points. Although historically points varied slightly by country, all points were approximately the same size. The modern, standardized system has exactly 12 points in a pica and exactly 6 picas in an inch. As such, there are 72 points in one linear inch (Cavanaugh, 1995; McLean, 1997). Comparing the relative sizes of type, however, can be most difficult: A typeface's size is not specified by the height of the capital letters. Rather, the size of a font refers to the maximum possible letter size. Because some typefaces have letters which are taller than the capital letters and many letters which have lowercase descenders which dip below the height of the capital letter, the capital height represents only part of the total size (see Figure 3). For example, Times is a smaller typeface than Helvetica even at the same point size because the descenders of Times are larger (McLean, 1997).

Figure 3
Perceptual size varies from point size

Even at the same point size, fonts appear smaller or larger. The largest determining factor in perceptual size is the ratio of the body to the ascenders and descenders.

Cavanaugh suggests that 12 point type "looks completely huge" in print and that typographers should start with a set of 10 point and increase or decrease as appropriate based upon the appearance when printed out (Cavanaugh, 1995, p. 106).

Print Research on Type Size

Paterson and Tinker did extensive research on how type size affects the legibility of type. Although specific results vary, the differences in reading rates at sizes from 6 points to 14 points were relatively small but highly statistically significant (Donald Gilbert Paterson & Tinker, 1929; Donald Gildersleeve Paterson & Tinker, 1932, 1940a, 1946; Tinker, 1963; Tinker & Paterson, 1929, 1943, 1944). For instance, one study indicated that 10 point size type is read faster than other sizes between 6 points and 14 points, being read at between 4.9% and 6.4% slower (Donald Gilbert Paterson & Tinker, 1929). A similar study with a different typeface showed that 9 and 11 point type size was read up to 2.3% faster than 10 point type (Donald Gildersleeve Paterson & Tinker, 1940a).

In a study of font selection and sizes of newspapers, Tinker and Paterson (1944) found that the most common sizes were 7, 7 1/2 and 8 points for type on the front page. Eight point type was used more often in 1942 than in 1935. Even in 1942, however, seven points was the most common size. Paterson and Tinker also compared the speed of reading of the most common newspaper sizes with the larger 10 point printing common in books of that era. Although standard newspaper printing was approximately 5% slower to read than the ideal size (Donald Gildersleeve Paterson & Tinker, 1946), printing smaller type equals less paper consumption. The inconvenience of a slightly slower reading speed is presumably offset by reductions in printing and distribution costs.

Tinker (1963) also studied user preference of type size. 225 readers considered the 11 point size most legible in print. Both 10 point and 12 point scored slightly lower with 8 and 9 point type receiving lower rankings.

Common Practice in Onscreen Type Size

As discussed previously and shown in Table 1, font size varies dramatically based on the client's choice of software. Note that it is not uncommon for type to be rendered much larger in one browser than another browser when the default settings are used. On Macintosh, for example, type rendered in Internet Explorer 5 is 45% larger than the same type rendered in Netscape 4.7.

Similarly, Adobe Acrobat often sizes the type to different pixel sizes based on the pixel resolution available and settings made in the PDF file. Early Macintosh monitors were designed to have exactly 72 pixels per inch. This afforded an exact match in physical size between screen and in print although the overall resolution (quality) was lacking. The choice of 72 ppi (pixels per inch) was deliberately set to match the design standard of 72 points per inch. Current monitors usually employ multisync technology. Multisync monitors are technically designed to support a specific number of horizontal and vertical pixels; however, these monitors will also support higher and/or lower resolutions (Apple, 2000a, 2000c). Because a user can set the logical resolution, it is difficult to determine the exact size of a pixel without knowing the attributes of the user's video card, the user's software settings and the hardware settings on each monitor (see Table 2).

Most software is designed to work best at 72 ppi. In multimedia applications, for instance, 72 ppi maps a graphic to each pixel on the screen regardless of dot pitch (actual size of each dot onscreen) (Apple, 1991, 1998; Kelsey, 2000). Microsoft software including Internet Explorer, PowerPoint and Windows assume that all monitors are 96 ppi (Armstrong, Herbert, & Gowin, 2000). In some Microsoft applications, it is possible to set the exact ppi and some attributes of the software will adapt to those settings (Vosseller, 2000).

Table 2
Sample Monitor Resolutions

	horizontal pixels	vertical pixels	diagonal size	mean ppi
Apple 15" LCD	1024	768	14.94"	85.69
PowerBook Pismo LCD	1024	768	14.06"	91.02
Sharp XGA Projector	1024	768	72.00"	17.78
AppleSync 1710	832	624	15.44"	67.37

Multisync monitors vary in pixels per inch based on technical attributes and user settings.

Onscreen Research on Type Size

Geske's (2000) "Readability of body text in computer mediated communication" is an experimental study of typeface size onscreen. Geske hypothesized that 14 point type would be more legible than 12 point type, which would in turn be faster than 10 point type. Geske's study also looked at font selection, choosing Palatino as a serif face and Helvetica as a sans-serif face. Geske appears to have used content and comprehension items developed specifically for this experiment. He reports that most selections included approximately 225 words and had a reading grade level of 7.5 in order to be appropriate

for his college age participants. Testing was completed using world wide web (web) technology. Line lengths were standardized at a length equal to the lowercase alphabet typed 2.5 times. Type was black on the default gray of the web browser. The experiment reports an N of 78.

The speed of reading findings surprised Geske. Although Geske (2000) predicted that speed would be higher at larger sizes, the findings did not support this. Of the six directional tests regarding reading speed, statistical significance was found only once when comparing 12 point type with serifs to 10 point type with serifs: 12 point Palatino (74.0 seconds) was read more quickly than 10 point Palatino (83.9 seconds).

The test of comprehension was equally remarkable. For the sans-serif typeface, the comparison between the 12 point size and 10 point size did result in greater recall. For the serif face, 12 point type resulted in greater recall than either 14 point or 10 point. The average serif comprehension score was 4.42 for 14 point, 4.85 for 12 point, and 3.98 for 10 point (Geske, 2000).

Rendering technology

When creating a book, electronic publishing computer files are processed and converted to high-resolution, orthochromatic negative film. This film, after being photographically developed and fixed, is used to make aluminum plates which are then placed onto a printing press. Because of the high-resolution available in negative film, book typefaces are rendered at resolutions of 2,400 orthochromatic pixels per inch or higher (Drewry, 2001). Even desktop laser printers use resolutions between 300 and 600 pixels per inch (Grant & Branch, 2000). Unfortunately, computer displays lack these

resolutions. Most computer displays offer between 72 pixels per inch and 100 pixel per inch resolutions (Geske, 2000; Vosseller, 2000).

Because the resolution of computer screens is so much lower than print forms, multimedia type is harder to read than print equivalents (Mason, 1997; Williams, 1994). Although work is underway to dramatically improve screen resolutions, these advances will not be widely available or affordable for years to come (Electronic Buyer's News, 1999; IBM, 2000).

Microcomputer Type Technologies

Only recently have technologies been in place to improve the rendering of type on screen. For most of the history of microcomputers, insufficient speed prohibited dedicating computer cycles to improving the appearance of type on screen.

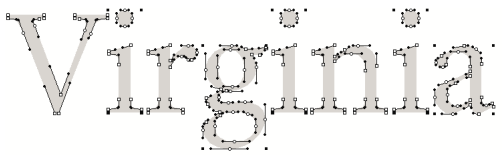
Early personal computers (PCs) simply stamped characters of a particular height and width onto the computer screen. Like a typewriter each character was mono-spaced. As such, there were a certain number of rows and columns available on the screen, usually only in one, primitive face.

With the entrance of mainstream graphical interfaces in 1984 with the Macintosh Lisa and Macintosh Plus, screen display of type was improved. A variety of distinct typefaces were available, many of them with proportional spacing. These bitmap fonts were still "stamped" onto the computer screen but the placement, size and choice of typeface was now within the user control.

Other typeface technologies were available for printing. Postscript fonts, a technology created by Adobe, could be used to print high-resolution vector fonts (see

Figure 4) to laser printers and high-resolution imagesetters. This created an environment where type could be specified on screen at exceedingly low quality and output at professional quality.

Figure 4
Many fonts are mathematically based



Professional typefaces are created using bezier vectors. These mathematical equations define each letterform using infinite resolution curves. Most typefaces now use this method to define these shapes. Vector fonts print at the resolution of the output device.

In 1990, Adobe introduced Adobe Type Manager (ATM). ATM allowed for the Postscript vectors from downloadable Postscript type to be used on screen. The real advantage of ATM was that it could create typefaces of minimal quality at any size, not just the sizes the bitmap manufacturer wanted. Unfortunately, at $150 to $300 per typeface only graphic and printing professionals used ATM and Postscript fonts.

Apple and Microsoft created a competing technology called TrueType. TrueType was designed for use on screen. The render engine was free unlike early versions of ATM and fonts were very inexpensive. Many fonts are available for free and libraries of hundreds of TrueType fonts were available for less than $100.

Although consumers readily accepted TrueType, graphic professionals shunned the technology because it did not scale up to professional grade equipment. Many experts claim that the underlying technology of TrueType was superior to Postscript (Microsoft, 1997). However, the professional type houses continued to develop for the more lucrative

Postscript market. As of this writing, there is little doubt that Postscript fonts are of higher quality and are far more reliable in commercial printing (Drewry, 2001).

Orthochromatic Type

Each pixel of a high-resolution print can have only two opposite values: black or white (in some cases ink or no ink, transparent or opaque). This process of only having two opposite states is known as orthochromatic mode. Orthochromatic mode is ideal for accurately defining letterforms and for representing pictures with a halftone. The low resolution of computer screens, however, poses a challenge in accurately displaying type (see Figure 5 and Figure 6).

Figure 5
Othrochromatic rendering of type

The word “Virginia” rendered as orthochromatic text in 14 point Adobe Minion. This is the way most computer programs render text today. Shown at 72 ppi at 400% scale.

Figure 6
Vectors not an exact match to raster grid

Because the resolution of the computer screen is very low, the render engine attempts to approximate the complex curves.

Anti-Alias Technologies

Fortunately, there is a way to minimize the impact of low resolution because computer monitors have the ability to display grayscale and color values. Each pixel on a computer monitor can have a variable tone. This allows for type and graphics to be anti-aliased. Instead of having harsh edges, grayscale can be used for softer transitions (Phong, 2000). Anti-alias graphics and text will be softer than their orthochromatic equivalent (Every, 1999; Landweber, 2000). Since these images contain more information, they may be more legible. Landweber (2000, p. 1) suggests that his anti-alias technique "effectively [doubles] your screen resolution."

To create an anti-alias image, the computer renders type to an off-screen buffer. Instead of creating this image at 72 ppi (pixels per inch, actual size for most software), a temporary, buffered version is created at a higher dpi and then resampled to actual size (72 ppi). Many anti-alias techniques use 216 ppi (three times actual size, see Figure 7). This process combines a three pixels wide by three pixels tall grid of black and white pixels to make one grayscale pixel (see Figure 8). In effect, the smaller black and white dots are averaged to create normal sized gray dots (Landweber, 2000).

Figure 7
Othrochromatic rendering of type, 216 ppi

This version of the word is rendered at 216 ppi instead of the usual 72 ppi. Creating the text at three times the horizontal resolution and three times the vertical resolution is the first step to creating an anti-alias version.

Figure 8
Anti-alias type created by downsampling

The 216 ppi image is then “downsampled” back to 72 ppi. This converts the high-resolution black and white data into low-resolution grayscale data. This results in nine levels of grayscale and presumably better quality than the black and white only version.

Although these grayscale images have less location data than the high-resolution equivalent, they obviously contain more visual data in the grayscale image than the low-resolution, orthochromatic norm. The perceived resolution of anti-alias type should be higher than the orthochromatic form based on this additional visual data. While an orthochromatic rendering of type will contain either 0% black (white) or 100% black, the 3x anti-alias rendering of type described here (adjusted-complex anti-aliasing) allows for nine grayscale states (0%, 11.1%, 22.2%, 33.3%, 44.4%, 55.6%, 66.7%, 77.8%, 88.9% and 100%).

Applications that use anti-Alias technology.

Prior to 1990, anti-alias rendering was restricted to a few high-end workstations (Negroponte, 1994). The last few years, however, have seen an increase in the use of anti-alias rendering techniques on desktop systems.

Adobe Photoshop was first to make widespread use of anti-alias technology. Because Photoshop began with only raster features, it was faced with the same resolution-dependence that still affects monitors. Essentially, early versions of Photoshop included

only pixel-based tools. When text or shapes that appeared to be mathematically drawn were needed, they had to be created using the pixel-based (raster) metaphor. No matter how graphics, images and text appear on screen, they were all made up of continuous-tone style pixels (picture elements, medium sized square dots which contain color or tone). In Photoshop, anti-alias type was an option to offer continuous tone in type. This continuous tone rendering style was essentially a compromise between vector test (not yet available in Photoshop) and the low-resolution photographic modes Photoshop is best known for. Photoshop 6.0 now offers vector type in some of its formats but retains three distinct anti-alias options—in addition to orthochromatic type—for use by web developers, multimedia developers and programmers (Adobe, 2000b).

Adobe Type Manager (ATM) was the next application to offer anti-alias type. ATM is a global render engine for Postscript font technology. Without ATM, Postscript fonts may be used on sophisticated printers but not onscreen. Postscript fonts are completely scalable onscreen with either the free or commercial versions of ATM loaded. Using recent versions of ATM, type can be rendered from vectors into orthochromatic mode and with anti-alias edges (Adobe, 2000c).

Adobe Acrobat Reader also includes anti-alias technology. Reader was the first consumer application to offer anti-alias support. The anti-alias algorithm used by Reader is very aggressive. At normal amounts of magnification, Reader offers 16 balanced levels of gray with all non-white tones relatively equally represented. Text as small as six points is anti-aliased. Below six points, type is represented as Greek, small gray bars that represent text when it is to small to be read (Adobe, 2000a).

Apple's Mac OS 8 offers similar functionality (Maurer, 1998). However, the font smoothing option in the Mac OS renders TrueType technology instead of Postscript fonts. This feature is less useful to graphic professionals who rely exclusively on Postscript fonts (Cavanaugh, 1995) but is far more useful for mainstream users, such as Internet surfers (Apple, 1996). Microsoft added similar functionality to Windows as an add-on for Windows95 and Windows98 (Microsoft, 1997).

Mac OS X continues Apple's goal to make anti-alias a standard feature. The new "Quartz" rendering technology makes extensive use of anti-aliasing, transparency, PDF and OpenGL. Under Mac OS X all fonts of any type and all 2D vector objects automatically inherent anti-alias technology (Apple, 2000b). There is currently no option to turn off anti-aliasing for small point sizes, certain font technologies (TrueType only or Postscript only) or globally (Siracusa, 2000).

Expert opinion on anti-alias rendering.

Different experts have varying thoughts on the value, if any, of anti-alias rendering. It is clear, however, that many people have strong opinions on the participant. For instance, Nicholas Negroponte (1994) has long called for all text to be anti-aliased. In the article "Aliasing: the blind spot of the computer industry," he says:

> What puzzles me the most is that we seem to have educated an entire generation of computer scientists who don't fully understand this simple phenomenon, and we seem to have trained the public to take it for granted. Perhaps it's time to make [orthochromatic] graphics

> a violation of Occupational Safety and Hazards Administration minimum standards for display quality. Or, perhaps the Environmental Protection Agency can declare this condition to be visual pollution. The point is that it must stop. (1994, p. 1)

Bayley (2000) agrees. He states that the anti-alias in Mac OS X (DP 4.2) is addicting and "painful to go back" (p. 1). The November 13, 2000, issue of the Independent takes the opposite extreme:

> It looked like someone had smudged the screen with margarine. Other people can bear [anti-aliasing], but I hate it; only by choosing a tiny size of a non-aliased font could I begin to write without feeling uncomfortable (Independent, 2000, p. 1).

Felici also thinks that anti-alias technology is counter-productive. Felici believes that anti-alias is fine for 14 point sizes and larger but that at the common text sizes of 10 points and 12 points, anti-alias is unacceptably blurry. Felici supports the use of orthochromatic bitmaps as the default reading onscreen technology. In fact, Felici (1996) thinks that hand-crafted bitmaps outperform any rendering technology.

Other sources have more ambivalent opinions. Colaiuta (2000), for example, thinks that anti-aliasing is an improvement, but that it will not be welcomed by all. Siracusa finds the quality reasonable and better than previous attempts but thinks that current

machines are not fast enough to anti-alias without disrupting the user-experience (Siracusa, 2000).

Web developers often use anti-alias in their designs. Because there are only a small number of fonts common to all internet-capable machines, rendering type into graphics is a common way to control the appearance of type. Programs like Photoshop can convert fonts into GIF and JPEG graphics thereby preserving the appearance of text as it appears on the designer's machine. Since different browsers and operating systems display type differently, rendering type is an easy way to control the exact pixel size of type. Some practitioners agree that type becomes unreadable onscreen at approximately 8 points (Fields, 2000; Gardner, 2000; Gowen, 2000). In fact, they state that anti-alias will either always or sometimes make type readable at sizes smaller than orthochromatic rendering (Fields, 2000; Gardner, 2000; Gowen, 2000). This is consistent with results on many web sites who use type as small as 6 or 7 points when anti-aliased (Cisco, 2000; XY_Art, 2000).

Research on anti-alias.

In addition to researching font selection, Weisenmiller (1999) compared the speed of reading and comprehension between orthochromatic presentation of type onscreen, anti-alias presentation of type onscreen and 600 dpi laser output. Weisenmiller found that there were statistical significant differences between the three presentation styles in both speed of reading and comprehension. As can be seen from the descriptive statistics in Table 3, anti-alias text outperformed 600 dpi print output which was only slightly better than orthochromatic text onscreen. Post hoc testing shows that the distinction regarding

speed of reading of either anti-alias presentation or print output compared to orthochromatic type was statistically significant; however, comparisons between anti-alias and print output were not significant. Table 4 shows a slightly different descriptive trend for comprehension. While the order of performance remains the same, print output was similar in speed of reading to anti-alias presentation than orthochromatic presentation. Once again, a Tukey test indicates that orthochromatic type shows a statistically significant difference in comprehension when compared to anti-alias onscreen type or print; but anti-alias and paper are not distinctly different (Weisenmiller, 1999).

Table 3
Weisenmiller's Speed of Reading Results

	M	*SD*
Orthochromatic	197.00	61.86
Anti-alias	221.93	67.47
600 dpi Print	201.14	54.42

Weisenmiller found statistical differences in speed of reading (words per minute) across the three presentation mode ($p < .05$). Weisenmiller also reports that Tukey post-hoc tests indicates that reading anti-alias type onscreen is faster than orthochromatic type onscreen and that reading 600 dpi print is faster than orthochromatic type onscreen. $N = 88$.

Table 4
Weisenmiller's Comprehension Results

	M	*SD*
Orthochromatic	23.705	6.721
Anti-alias	27.125	6.475
600 dpi Print	26.716	6.621

Weisenmiller found statistical differences in comprehension across the three presentation modes ($p < .05$). Weisenmiller also reports that a Tukey post-hoc test indicates that reading anti-alias type onscreen results in better recall than orthochromatic type onscreen; another Tukey test shows that reading 600 dpi print results in greater recall than reading orthochromatic type onscreen. $N = 88$.

Secondary Variables

There are a variety of other variables that may affect the results of experimentation into the onscreen presentation of type. While these variables are not manipulated in this study, these factors must be carefully controlled to maintain validity and maximize reliability.

Environmental Lighting in Print Research

There is an entire body of research dealing with light intensity and its associated impact on reading. In general, no one type or color of light is superior to any other when reading text on paper (Tinker, 1939). As long as there is adequate illumination, reading rate is not affected. Lighting levels from 25-foot candles to 100 foot candles result in no change in speed of reading (Tinker, 1959; Weston, 1935, 1945). While no other attributes are likely to overcome too little light or too much glare, illumination and contrast play very little part in outcomes regarding speed of reading (Weston, 1935, 1945).

Angle of Reading in Print

There are several studies that deal with the best viewing angle for reading books. It has been consistently found that a book slanted at a 45° angle, as might be experienced on a lectern results in maximum reading speed and therefore offers the best legibility. As the angle deviates from this angle, reading speed is hampered. For a book held vertical, the speed of reading has been shown to be slowed by 5.7% (Tinker, 1956). Placing a book flat on a table has been shown to slow speed of reading by 9.8% (Skordahl, 1958). It is unclear how the literature on book reading being optimized at a 45° angle relates to computer screens, which are generally vertical.

Line Length in Print

Typographers on Line Length.

Typographers have long attempted to find a comfortable line length. Traditional wisdom suggests that the shortest line still easy to read is 39 characters long, the equivalent of 1.5 lowercase alphabets.

Bringhurst (1999) suggests that any number of characters between 45 and 75 characters is appropriate with 66 characters considered the best line length. Applying Bringhurst's basic approach to average typeface width suggests lines of 20 to 40 times the point size are acceptable. This approach suggests that the line length should be scaled to the size and shape of the letterforms of the specific typeface. An application to 10 point Times results in an ideal measure (measure is the typographic term for line length) of roughly 23 picas.

Cavanaugh makes similar recommendations regarding line lengths. His ideal line length for 10 point times is between 18 and 30 picas (3.0 to 5.0 inches). The low end of this range is slightly narrower than Bringhurst while his maximum length is close to Bringhurst's ideal measure. Cavanaugh (Cavanaugh, 1995) suggests that the primary benefit of shorter lines is increased white space.

Research on Line Length.

There is extensive literature dealing with line length in print. This literature confirms best typographic practice (Luckiesh & Moss, 1941; Donald Gildersleeve Paterson & Tinker, 1940a, 1940b, 1942; Starch, 1923; Tinker & Paterson, 1929). Although there are slight variations within the treatments and results of each study, clear patterns emerge. Type set at 10 points solid (no additional vertical space) reads fastest around 19 picas (3.16 inches). Setting lines with normal lead (a 20% increase in vertical spacing), as is suggested by fine typography (Cavanaugh, 1995), results in fastest reads at lengths up to 31 picas (5.16 inches). Type set at 12 points with normal lead read equally well from 25 picas (4.16 inches) to 33 picas (5.5 inches). There appears to be a fairly broad range of values around each best width that reads either as fast or very close to as fast as the ideal. As type size is reduced, the line length for fastest reading is also scaled down.

Legibility, Speed of Reading and Comprehension

Introduction to Legibility and Speed of Reading

According to Miles Tinker, a variety of tests have been conducted to determine the legibility of type. Some of these include speed of perception, perceptibility at a distance,

perceptibility in peripheral vision, Luckiesh-Moss visibility threshold, reflex blink technique (more blinking equals less legibility), rate of work, eye movement, and fatigue in reading (Pearson, Barr, Kamil, & Mosenthal, 1984; Tinker, 1963). Tinker (1963) and Sutherland (1989) suggest that three speed of reading tests—Chapman-Cook Speed of Reading, Tinker Speed of Reading, and Minnesota Speed of Reading—have been the most used in print research.

Speed of Reading Tests

Several variations on these speed of reading tests were widely used from 1924 until 1963. The seminal Chapman-Cook Speed of Reading Test provided a basis for later tests including the Tinker Speed of Reading Test and the Minnesota Speed of Reading Test. All of these tests used short statements that were not internally consistent. In the Chapman-Cook and Tinker tests, participants were asked to identify the word that should be changed to correct the inconsistency. All of these test problems were designed so that very little time was spent thinking about the question. As such, identification scores were very high. Because the processing was simple, the cognitive component is minimized and the amount of time spent on each problem reveals the relative legibility of the test itself.

Chapman-Cook Speed of Reading Test.

The instruction page of one typographic variant (the six printing unit arrangement) of the Chapman-Cook test is provided in Appendix A (Chapman, 1923; Donald Gildersleeve Paterson & Tinker, 1940a). In this version of the test items are formatted similar to paragraphs. Making some items continuous in copy while providing a

paragraph space between every few items was intended to make the legibility of type in the test similar to legibility in a typical reading assignment.

The Chapman-Cook test recently has been revived for use in brain research. Using a conventional formatting of the test, doctors in the Department of Neurology at the University of Iowa College of Medicine have found that a time sensitive test is a much better indicator of brain trauma than more traditional comprehension tests (Manzel & Tranel, 1995). In essence, this research confirms the claim that speed of reading tests are not only useful tests but may also reveal smaller distinctions in speed of reading than comprehension tests. As previously mentioned, this finding is contradicted by Geske's recent study (Geske, 2000) which found more effects in comprehension than speed of reading.

A copy of the Chapman-Cook test has been obtained and can be found in Appendix B. The test is made up of 25 short items and is typically completed in 2 1/2 minutes for elementary audiences and 1 3/4 minutes for college audiences (Donald Gildersleeve Paterson & Tinker, 1940a).

Tinker Speed of Reading Test.

The Tinker Speed of Reading test was designed to be a longer form of the Chapman-Cook test. It uses the same basic format and approach as the Chapman-Cook and Minnesota tests. This example has a word or phrase in the second sentence that invalidates the greater meaning of the passage. Try question 12.

12. A certain doctor living in a city near here always had a very serious expression on his face. This is perhaps because in his

work he meets only well people. (Tinker, 1963)

In this item, from the Tinker test, the complete example as written doesn't make sense. However, by changing the word "well" in the second sentence to "sick" the item becomes internally consistent. Students were asked to identify the wrong word or phrase and cross it out with a pencil. The test was scored for accuracy but also for how many items the student was able to complete in a short period of time. (Tinker, 1963)

Length of test was found not to be significant. Tinker suggest that tests as short as 90 seconds offer a valid measure of legibility (Tinker, 1963). Other sources suggest that 60 seconds may be adequate (Pyke, 1926; Weston, 1935, 1945).

The Tinker Speed of Reading test has not been obtained; extensive research by multiple parties indicates that only one copy of the test is likely to be accessible. This copy is part of the Southern Illinois University library's Historical Test Collection (Person, 2000) and may be viewed in person for research purposes.

Minnesota Speed of Reading Test.

The Minnesota Speed of Reading Test was similar to the Chapman-Cook and Tinker test but was designed specifically for college students. The Minnesota Speed of Reading Test asks the participant to identify a phrase that conflicts with the greater meaning of a short passage. The Minnesota test comes in two forms of 38 questions each (Eurich, 1936). A copy of the Minnesota test has been obtained. It is Appendix C.

Statistics of speed of reading tests.

Tinker claims that when used in realistic practice speed of reading offers "one of the more satisfactory methods of investigating legibility of print" (Tinker, 1963, p. 23). For the Chapman-Cook Speed of Reading test, reliability between the two forms is r = .86 when work-limit is held constant and r = .84 when time-limit is held constant (Tinker, 1963).

Test of Speed of Reading and Comprehension

Nelson-Denny Test.

The Nelson-Denny Reading Test is a reading test that has been revised several times over the last 40 years. The test includes forms to test both vocabulary and comprehension. The test is generally timed (Dyer, 1997; Riverside-Publishing, 2000). The test has been widely used as a test of comprehension and speed in print and onscreen (Canary, 1983; Dyer, 1997; Holmes, 1986; Journa, 1989; Taylor, 1990; Tullis, Boynton, & Hersh, 1995; Turner, 1982; Weisenmiller, 1999). To determine speed of reading, participants read for a specific time limit. At the end of the allotted time, students note the last word read. To determine comprehension students read a short passage and then fill in blanks every few words to demonstrate comprehension.

GRE Practice Test of Comprehension and Speed of Reading.

While the Graduate Record Exam reading test is traditionally thought of as a test of comprehension, a recent study by Grant and Branch (2000) collected data on both speed of reading and comprehension. This study compares serif and sans-serif typeface for

testing on the world wide web. By having only the passage from the 1997 practice test visible until a user clicks, they collected data on how long it took for each participant to read the passage.

Summary

Hundreds of studies have looked at the legibility of type and learning that results from differing typographic factors. As we move towards a more digitally based world, reading onscreen for education, business and entertainment will likely continue to increase. There is a small but growing body of work that looks how best to present words onscreen.

This study experimentally tests several key variables while controlling for other factors. Through research of this kind, speed of reading and comprehension can be maximized for this new area in educational research.

Need for the Study

One should be careful to note the methodologies employed in previous research. Isaacs states that much of the research in this area is “speculative” (Isaacs, 1987). In fact, little of the literature regarding type design onscreen consulted for this review of literature was experimental in nature. Perhaps the current thinking about type design is based on historical preference, out of date information or conjecture. The lack of experimentally supported findings regarding multiple variables should underscore the need for this type of research.

As early as 1931, Buckingham argued that much research regarding type in print was based on "imagination." He noted that there are significant methodological flaws in the research and that investigators need a background in typography before undertaking this type of research. He also commented that the univariate model is particularly suspect and argues that multiple variables be examined simultaneously to more accurately understand the relationships between them (Buckingham, 1931). Another more recent critique of typographic research suggests that research needs to be as close to real-world conditions as possible and as similar to tasks completed by practitioners (Hartley & Burnhill, 1977).

Moore (1993) raised similar concerns about multimedia in general in the article "Multimedia: Promise, Reality & Future." Making the claim that the "most prevalent sources… are assumption, intuition, and (apparently) commonsense." Moore argues that not only should more research be undertaken but that new research should be methodologically sound. This study addressed precisely that concern.

Vast resources are being focused on multimedia-based education. The prevalence of Internet instruction and interactive multimedia is reshaping the face of education and commerce. And yet, very little is known about how onscreen type is perceived by the learner. Experiments, such as this one, are critical to exploring the role and importance of type design in learning through quantifiable, repeatable experimentation.

Use of the world wide web and other online resources has grown as well. Nielsen/NetRatings report that the average home user spends 20 minutes per day online. Access to the Internet from home is also increasing with 54% of all Americans having access from home (Mariano, 2000). A simple calculation shows that Americans are spending approximately 50 million hours each day in front of computer screens at home

and far more time viewing computer generated type when computer use at work is included.

Similarly, there is a need to provide evidence in support of or against the handful of previous onscreen studies. Plus, this study considers an under-investigated variable. This study is a step towards understanding the role of type in modern interactive instruction and in maximizing the outcomes and effectiveness of multimedia instruction.

Hypotheses

The proposed study seeks to explore the impact—if any—that rendering technology, font selection and type size has on student comprehension and time to complete an educational task. Specifically, this study addresses testing delivered on a computer screen using interactive multimedia.

Research Questions

H1) Does the use of serif type vs. sans-serif type result in differences in reading speed (words per minute)?

H2) Does the use of serif type vs. sans-serif type result in differences in reading comprehension?

H3) Does the choice of type size (8 point, 10 point, 12 point sizes) result in differences in reading speed (words per minute)?

H4) Does the choice of type size (8 point, 10 point, 12 point sizes) result in differences in reading comprehension?

H5) Does the choice of type rendering style (orthochromatic vs. anti-alias) result in differences in reading speed (words per minute)?

H6) Does the choice of type rendering style (orthochromatic vs. anti-alias) result in differences in reading comprehension?

H7) Is there an interaction between or among type selection, type size and type rendering style with regard to speed of reading?

H8) Is there an interaction between or among type selection, type size and type rendering style with regard to reading comprehension?

Variable List

Independent	Type Rendering Technology
Independent	Font Selection
Independent	Type Size
Dependent	Legibility
Dependent	Comprehension

Variables of Interest

Independent Variables

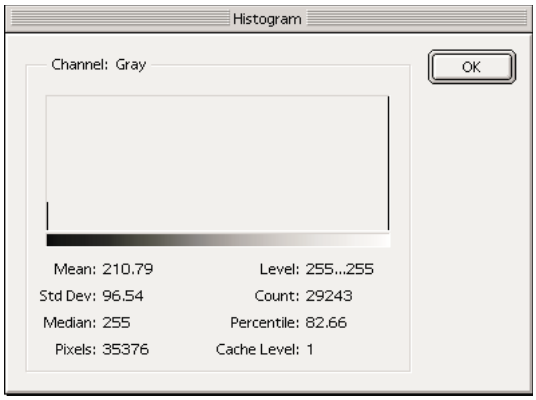
Type rendering technology.

As previously discussed, onscreen type is rendered differently in different contexts. Although most programs inherit rendering technologies from the operating system, a few applications handle text rendering with a custom algorithm (Adobe, 2000a; Adobe, 2000b; Adobe, 2000c; Apple, 2000b; Maurer, 1998). While most Macintosh applications simply use the built in APIs (application program interfaces) for rendering text, a few use their own technologies. For instance, Adobe Acrobat uses a distinctly soft-looking technique that balances 16 levels of gray instead of the more heavily weighted tonal ranges. Adobe has also extended their type scaling engine, called Adobe Type Manager (ATM), to include anti-alias technology. ATM is available as a replacement or hook for the Apple rendering engine. Another unique rendering engine is built into Adobe Photoshop. Because of the raster nature of text exported from Photoshop, it offers multiple unique options to control rendering style.

As can be seen from the figures 9–13 demonstrating and describing the tonality of different rendering technologies, all of the anti-alias technologies are similar in terms of overall appearance and in terms of use of the tonal range.

Figure 9
Photoshop's "none" type style (orthochromatic)

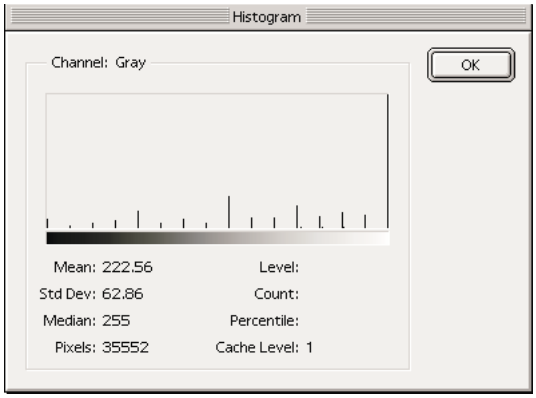
Psalm 23: A Psalm of David
The Lord is my sheperd; I shall not want. He maketh me to lie down in green pastures; he leadeth me beside the still waters. He restoreth my soul: he leadeth me in the paths of righteousness for his name's sake. Yea, though I walk through the valley of the shadow of death, I will fear no evil: for thou art with me; thy rod and thy staff they comfort me. Thou preparest a table before me in the presence of mine enemies: thou anointest my head with oil; my cup runneth over. Surely goodness and mercy shall follow me all the days of my life: and I will dwell in the house of the Lord for ever.



Photoshop's none style renders the type with no tone. Each pixel of the text is either completely on (black) or completely off (white). The histogram provides a visual and mathematical explanation of the tonality of the image.

Figure 10
Photoshop's "crisp" type style (anti-alias)

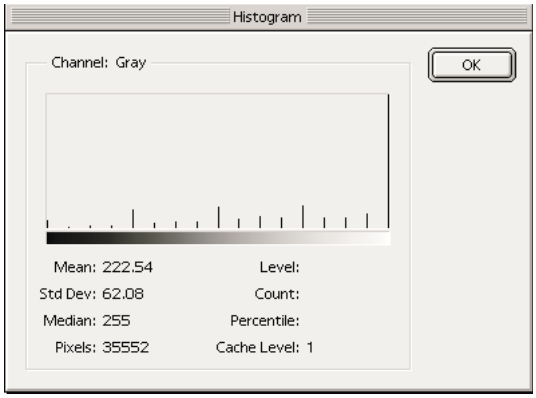
Psalm 23: A Psalm of David
The Lord is my sheperd; I shall not want. He maketh me to lie down in green pastures; he leadeth me beside the still waters. He restoreth my soul: he leadeth me in the paths of righteousness for his name's sake. Yea, though I walk through the valley of the shadow of death, I will fear no evil: for thou art with me; thy rod and thy staff they comfort me. Thou preparest a table before me in the presence of mine enemies: thou anointest my head with oil; my cup runneth over. Surely goodness and mercy shall follow me all the days of my life: and I will dwell in the house of the Lord for ever.



Photoshop's crisp style renders the type with an anti-alias. Using 16 levels of gray, the crisp style is typical of modern anti-alias techniques.

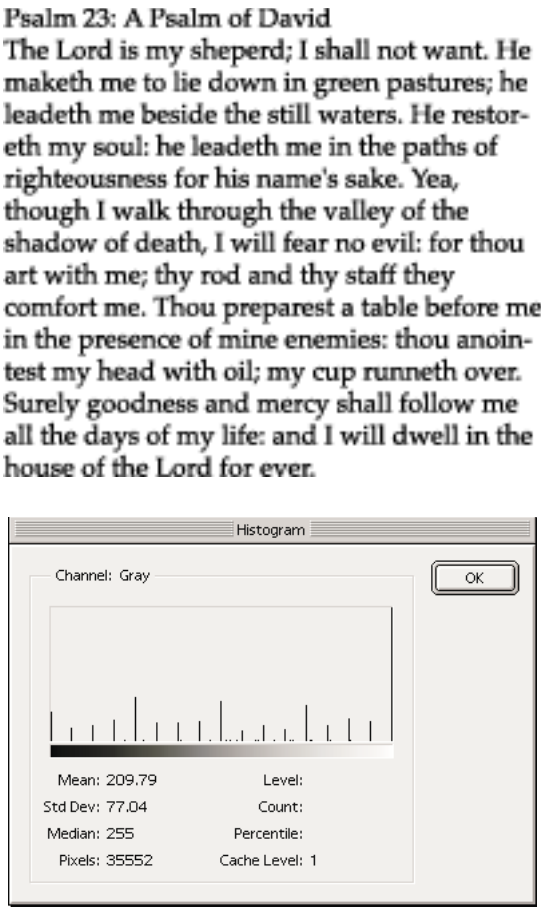
Figure 11
Photoshop's "smooth" type style (anti-alias)

Psalm 23: A Psalm of David
The Lord is my sheperd; I shall not want. He maketh me to lie down in green pastures; he leadeth me beside the still waters. He restoreth my soul: he leadeth me in the paths of righteousness for his name's sake. Yea, though I walk through the valley of the shadow of death, I will fear no evil: for thou art with me; thy rod and thy staff they comfort me. Thou preparest a table before me in the presence of mine enemies: thou anointest my head with oil; my cup runneth over. Surely goodness and mercy shall follow me all the days of my life: and I will dwell in the house of the Lord for ever.



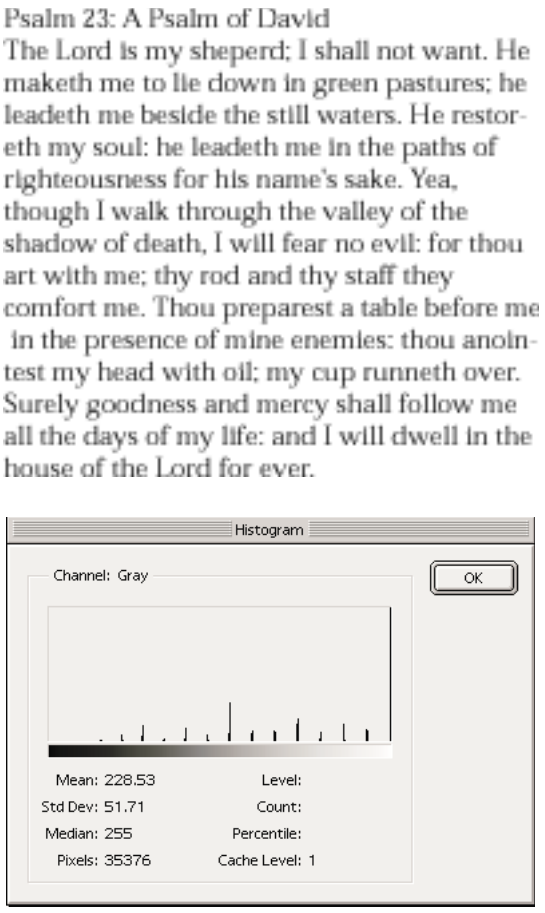
Photoshop's smooth style renders the type with an anti-alias. Using 17 levels of gray, the smooth style has a soft, fuzzy look. Use of smooth anti-alias techniques have grown more common as Moore's Law increases processing speeds.

Figure 12
Photoshop's "strong" type style (anti-alias)



Photoshop's strong style renders the type with an anti-alias. Using 17 levels of gray, the overall effect is quite dark.

Figure 13
Acrobat type rendering (anti-alias)



Acrobat uses a custom anti-alias rendering technology. The technique appears quite soft, in large part because the tonal range is well balanced with very few black or dark gray pixels.

The application used usually plays no roll in the rendering of type. In general the applications are simply a vehicle for the delivery of type created in one of many similar rendering-technology engines. Most applications simply inherit the typographic features of the operating system or of a custom type rendering utility. Only Adobe Photoshop and Adobe Acrobat disregard the traditional type rendering APIs (Adobe, 2000a, 2000b,

2000c; Apple, 2000b; Maurer, 1998). As such, rendering and delivery applications are less important than the selection of a render engine.

Photoshop's "crisp" rendering style is the appropriate representative of the anti-alias style because it offers a typical tonal range (see figures 9-14) and is consistent with type creation tools used by many creators of educational multimedia (Williams, 1998; Phong, 2001). Photoshop's "none" style was used to represent orthochromatic type primarily because it could be created and manipulated with the same tools and formatting used for the "crisp" rendering style.

Font selection (sans-serif vs. serif).

Geske (2000) used Palatino and Helvetica for his recent study although it is unclear whether he considered font file format. The font file format in his experiment was most likely TrueType since that is the file format built into the Macintosh operating system. Grant (2000) used Helvetica as one of his two sans-serif fonts and variants of Times for serif type. However, Grant made little effort to control for the specific typeface, computer architecture, size, or experimental environment.

Type size.

Typographic research in print has considered a wide range of type sizes appropriate for use as body copy. A survey of newspapers, for instance, found sizes from 6.5 points to 12 points used (Tinker & Paterson, 1944). Onscreen studies have tended to use larger sizes, presumably because of the lower resolution of this medium: Geske (2000) used 10

point, 12 point and 14 point sizes. And although, Geske thought larger sizes would be read more quickly, the results did not confirm this hypothesis.

In some cases it would be appropriate to resize one face to match the other size. Because the size of a typeface is based not only on the size of the uppercase letters but also the ascenders and descenders on lowercase letters, the body size of the typefaces should be compared. To perfectly match Palatino and Helvetica, Helvetica would have to be scaled down slightly to 97.75%. However, because the resolution onscreen is so low, this difference is substantially less than one pixel. As such, this study did not scale either font and accepted the typefaces as being the same relative size when used onscreen.

Using smaller font sizes was thought to likely increase the impact of other variables. The smaller sizes of type may degrade speed of reading and comprehension, therefore making the impact of the other variables more observable. While a variable such as anti-aliasing may provide only a small impact on larger sizes of type that are already quite legible, anti-aliasing may demonstrate a dramatic interaction at small sizes.

Dependent Variables

Speed of reading.

To test for speed of reading, it is customary to use a test that requires very little cognitive processing. When the activity is simple, only speed of reading remains. The most common speed of reading test is the Chapman-Cook (Appendix B). This test has been very widely used in typographic research (Chapman, 1923; Paterson & Tinker, 1940a; Manzel & Tranel, 1995; Tinker 1963; Pyke, 1926; Weston, 1935; Weston, 1945). This proposal suggests the use of the Chapman-Cook Speed of Reading Test.

Comprehension.

In addition to observing the speed of reading, it is also important to know what the larger educational impact is. Comprehension is one common indicator of educational impact. One of the most commonly used and standardized tests of verbal comprehension is the GRE test of verbal comprehension (Appendix L).

Font File Format

Although this study did not manipulate the font file format, the issue of font file format must be carefully considered. There are three different font file formats that could have been used for an instrument of this type: bitmap, TrueType and Postscript. Other studies have not carefully considered the positives and negatives of each format. Bitmap fonts are hand-drawn raster versions of a font available only in specific, screen quality sizes. Because bitmap fonts are not available in an anti-alias version, they are not appropriate for this study. TrueType fonts can be rendered in both variations. However, many TrueType fonts include custom optimization for presentation onscreen. As such, using TrueType would introduce an uncontrolled variable that could influence the outcome of this experiment. Postscript fonts can be rendered onscreen in multiple modes and sizes. Because postscript fonts were originally intended for use in print, they do not include specific optimization when presented onscreen. As such the use of Postscript fonts onscreen creates an equal and level playing field to evaluate multiple variables; additionally, the use of non-optimized file format technology may make it easier to observe other main effects and interactions.

Methodology

Research Design

Two distinct 2 x 2 x 3 experimental designs were employed for main effects of type rendering technology (anti-alias, orthochromatic), type style (serif, sans-serif) and type size (small, medium, large) on legibility and comprehension. Both experiments were combined into one instrument with two distinct sections. This approach allowed for a study that could not only search for main effects but also for interactions between and among the independent measures. Earlier studies have been criticized for not considering multiple independent measures (Hartley, 1977; Buckingham, 1931). Font selection and type size will be manipulated between participants while type rendering technology will be manipulated within participants.

The electronic instrument derived largely from widely-used, standardized tests was administered to 117 student volunteers at a large land-grant university in southwestern Virginia.

The first part of the interactive multimedia instrument used questions from the Chapman-Cook speed of reading tests while the second part incorporated 16 comprehension items appropriate for college learners from the 1995–1996 Graduate Record Exam practice test. The instrument included 12 distinct presentation styles representing each combination of the independent variables. For example, one “instance” of the instrument offered serif type at medium size. Every participant saw half of the instrument in an anti-alias format and half of the instrument in orthochromatic format.

The instrument was developed for a specific class of Macintosh computer, the PowerMacintosh 7300 with Applevision 17" (16.1" viewable) monitors running at 832 x 624 pixels of resolution. These specifications were constant for all participants of the study. Different computer architectures and configurations can have significant variations in factors such as monitor size and dot pitch (the size of each pixel). To control for these variances, a standard configuration was adhered to for this study.

As seen in Table 2, the Applevision 1710 monitor has a 67.37 ppi resolution. The chairs were affixed to the floor to keep the participants pupils approximately 18 inches from the computer screen. The monitors were tilted slightly towards the participant to minimize the effect of reading angle. For most participants, eye level was approximately 25% from the top of the monitor's viewable area and 75% from the bottom of the viewable area.

The interactive program randomly assigned students to treatments. The students read the content at their own pace and progressed to a set of standardized comprehension questions. The students' responses were timed and graded by the program. Comprehension responses and time to complete each item were stored on a private area network server. Only a short numeric identifier linked participants with their data.

Operational Variables

Independent Variables

Type rendering technology (anti-alias vs. orthochromatic).

All text in the instrument was pre-rendered into raster graphics to control for the precise rendering of the text. A review of the technical literature indicates that most

software defers font rendering tasks to the operating system. Only Adobe Photoshop and Adobe Acrobat use custom rendering technologies (Adobe, 2000a; Adobe, 2000b; Adobe, 2000c; Apple, 2000b; Maurer, 1998). As such, rendering and delivery applications are less important than the selection of a render engine.

Photoshop's "crisp" rendering style was selected because it is the most representative of the anti-alias style as it offers a typical tonal range (see figures 9-14). Additionally, using a Photoshop anti-alias rendering style is consistent with type creation tools used by many creators of educational multimedia (Williams, 1998; Phong, 2001). Photoshop's "none" style was used to represent orthochromatic type primarily because it can be created and manipulated with the same tools and formatting used for the "crisp" rendering style. All type was created with fractional widths disabled.

Font selection (sans-serif vs. serif).

Another variable that was considered was font selection. Adobe Helvetica regular (non-bold, non-italic, medium weight) was used as the sans-serif typeface while Adobe Palatino Roman (non-bold, non-italic, medium weight) was used as the serif face. Both Helvetica and Palatino are commonly used typefaces that can be found on most common computer systems and are included in the installation software for most network printers.

Each font selected is a representative sample from the group of serif and sans-serif typefaces. Helvetica is a clean typeface with evenly weighted strokes. As such, it reproduces well onscreen and in print for headlines and body copy. To the contrary, many other sans-serif typefaces are much more irregular and stylized. These stylistic faces are more often used for special purposes and headline applications. Palatino is, likewise,

representative of the serif faces. Designed with classical lines influenced by the Renaissance and lettered with prominent serifs, it offers a clean, readable face. The prominent serifs are thought to increase readability in print (Smeijers, 1996); the impact of serifs onscreen is still a matter of dispute.

Type size (8, 10, and 12 points).

Earlier studies used slightly larger typeface sizes than this study. For example, Geske (2000) used 10, 12 and 14 point sizes for his experiment. Geske expected that larger sizes would be more legible–and therefore read faster–although his study did not support this finding.

Due to the differences in the average monitor resolution, fonts can and are routinely delivered in different pixel sizes on varying computer platforms. To control for these factors, this study was delivered entirely on Macintosh computers. A simple calculation based on the presumed resolution of each platform (96 ppi for Microsoft Windows, 72 ppi for Macintosh OS) shows that Macintosh fonts can be rendered at 75% pixel size and still match the physical size of a PC monitor. While the specific resolutions of each individual monitor are different, the concept that the pixel sizes are smaller on the Mac is valid.

The instrument used three sizes: 8 point, 10 point and 12 point sizes. These typefaces represent common sizes in use onscreen (Cisco, 2000; McLean, 1997; XY_Art, 2000).

Dependent Variables

Legibility: time spent on speed of reading test.

The first component of the instrument was a legibility test that was measured by speed of reading. An electronic version of the Chapman-Cook test was used to determine the speed of reading. This test asked participants to identify if the word in the second half of a sentence violates the greater meaning of the passage. Problem 7 is shown below. In this example, the word "lemons" violates the greater meaning of the question because lemons are not sweet. The entire Chapman-Cook instrument as used in earlier, print studies can be found in Appendix B.

7. Some people are fond of sweet things.
They put lots of sugar in their coffee, and
eat lemons for the same reason.
(Chapman, 1923)

Participants were shown two sample questions before starting the 24 question Chapman-Cook test. In this electronic implementation, students could spend as much time on each question as they needed. The length of time used for each question was recorded in 1/60th of a second intervals. Time for each question was aggregated and these totals were converted into words per second.

Comprehension: score on comprehension test.

The Graduate Record Exam (GRE) from the Educational Testing Service includes verbal, math, analytical and specialty sections. The instrument in this study used four passages and sixteen questions from the verbal comprehension section of the 1995-96 General Test Descriptive Booklet (Rozmiarek, Burgess, & Weinfeld, 1996). Students read a short passage and then proceeded to answer four questions about each passage. A sample question is shown below.

> Mars revolves around the Sun in 687 Earth days, which is equivalent to 23 Earth months. The axis of Mars' rotation is tipped at a 2.5 degree angle from the plane of its orbit, nearly the same as Earth's tilt of about 2.3 degrees. Because the tilt causes the seasons, we know that Mars goes through a year with four seasons just as the Earth does.
>
> From the Earth, we have long watched the effect of the seasons on Mars. In the Martian winter, in a given hemisphere, there is a polar ice cap. As the Martian spring comes to the Northern Hemisphere, for example, the north polar cap shrinks, and material in the planet's more temperate zones darkens. The surface of Mars is always mainly reddish with darker gray areas that, from the Earth, appear blue green. In the spring, the darker regions spread. Half a Martian year later, the same process happens in the Southern Hemisphere.
>
> One possible explanation for these changes is biological: Martian vegetation could be blooming or spreading in the spring. There are

other explanations, however. The theory that presently seems most reasonable is that each year during the Northern Hemisphere springtime, a dust storm starts, with winds that reach velocities as high as hundreds of kilometers per hour. Fine, light-colored dust is blown from slopes, exposing dark areas underneath. If the dust were composed of certain kinds of materials, such as limonite, the reddish color would be explained.

1. It can be inferred that one characteristic of limonite is its

(a) reddish color.

(b) blue-green color.

(c) ability to change color.

(d) ability to support rich vegetation.

(e) tendency to concentrate into a hard surface.

The second dependent measure was an aggregated score on a multiple-choice test. Students read these short passages of text in one of the twelve experimental, typographic styles being studied and then answered four comprehension questions in the same style about what is implied by the passage. The correctness of each answer was sent to and stored in the online database.

Participants

A total of 117 participants completed the legibility and comprehension instruments. Of these five were removed from the dataset for having uncorrected vision while taking the test and two were removed because they did not complete the entire test.

Participants were recruited from a variety of classes in the visual arts: Art History discussion (sophomore level course, required for all art majors, N=58), Topics in Graphic Design: Layout (junior level course, N=3), Electronic Prepress (sophomore level course, most students are juniors, N=27), and Professional Seminar (senior level course, N=22).

The sample was comprised of slightly fewer females than males; there were 51 women (43.4%) and 59 men (56.6%).

The mean date of birth reported was 1979.4. The test was administered in May 2001.

Although five participants were eliminated from the sample because their vision was uncorrected when they took the test, most had either normal or corrected vision. Out of 110 participants with normal or corrected vision, 63 (57.3%) reported normal vision while 47 (42.7%) reported corrected vision.

Participants were asked how much time they spent reading on paper and onscreen. Detailed descriptive statistics can be seen in Table 5. The data suggest that these participants are experienced computer users. In fact, a review of the quartile data shows that many participants spend as much time reading onscreen as in print. Further, the most serious computer users read substantially more onscreen than in print.

Table 5
Amount of Time Participants Spend Reading

	In Print	Onscreen	Surf Web & Download	Chat
Mean	70.62	105.02	61.60	56.25
Median	60.00	60.00	42.50	30.00
Std. Deviation	57.88	101.61	66.96	89.79
Minimum	2.00	5.00	0.00	0.00
Maximum	360.00	600.00	360.00	600.00
25 Percentile	30.00	30.00	15.00	0.00
50 Percentile	60.00	60.00	42.50	30.00
75 Percentile	97.50	122.50	90.00	60.00

Observe that many students spend roughly equal amounts of time reading online and offline. Some students, however, spend a very large amount of time online thereby skewing the means. *N* = 110.

Procedures

Overall description

Participants completed two distinct tests via a computer-delivered instrument. Up to five students in the same room completed the instrument simultaneously on locally networked computers. The participants completed both sections of the instrument using identical computers and identical monitors set to factory settings. Chairs had their movement restricted so that all participants were the same approximate distance from the monitors. All natural light was blocked with normal fluorescent lighting illuminating the area.

Students were advised of their voluntary participation and were asked to sign a consent form in keeping with the university's institutional review board policy. Students were asked to provide basic demographic information via an online data collection mechanism. Following this task, the students completed two test components: the first component asked them to identify the word or phrase that was not consistent with the

greater meaning of the sentence. For example, in question four, the word “hammers” does not belong and should actually be “saws.”

4. We started to cut down a tree in our
front yard, but after working for two and
a half hours, we gave it up, because our
hammers were no good.

The Chapman-Cook instrument was developed for use in print. Participants of the original version were asked to cross out the word that spoiled the greater meaning. The computerized version used in this study adapted the instrument for use as a computer-based, online test. Five words of similar length and complexity were identified from the second half of each sentence and used as potential multiple choice answers. When viewed onscreen, participants will see the original passage on the left of the screen and the five potential answers on the right of the screen. Each answer is preceded by a number to simplify student response and assist in electronic grading.

Student responses were recorded and graded. Comprehension in this section was expected to be very high because the task was cognitively simple. As such, speed of reading times for each participant was also collected. By comparing the speed of reading times between and among variations of the instruments, differences in speed of reading—and therefore legibility—can be statistically analyzed.

The second component was a more traditional test of reading and comprehension. After reading a short passage, students were asked a series of multiple choice questions to test how well they understood the passage.

Data regarding time and comprehension were collected by the instrument and stored on a local server. This comma-delimited data was verified and simplified in Excel with inferential data analysis performed in SPSS 10.

Assignment of Participants

Participants were assigned to the different treatments at random. The instrument included a random number generator and assigned participants to treatments when the instrument was started. Because of this, perfectly balanced groups were not obtained. Even without stratification of the sample, however, the samples in each treatment were relatively well balanced.

Treatment-by-treatment Procedures

The instruments in all cases were identical except for the typographic variables being manipulated. Font selection and type size were manipulated between participants while type rendering technology was manipulated within participants.

As such, all participants encountered both rendering technologies. For some participants the Photoshop “crisp” anti-alias came first in the Chapman-Cook component while for others the Photoshop “none” rendering technology appeared first. For the comprehension component, type rendering style was manipulated across problems (each problem is comprised of a passage and four related questions). Each participant

encountered two problems (10 screens) in the Photoshop "crisp" anti-alias and two problems (10 screens) in the Photoshop "none" rendering style.

Each participant was randomly assigned to one of the six remaining style categories: a) Palatino, 8 points; b) Palatino, 10 points; c) Palatino, 12 points; d) Helvetica, 8 points; e) Helvetica, 10 points; f) Helvetica, 12 points. These six baseline styles were used in the instrument for data collection, instructions and content.

Materials

Standardized Instruction

The electronic instrument was comprised of screens which collected demographic information, screens which provided the Chapman-Cook instrument, and test screens which provided the GRE comprehension items. Screen captures from the instrument can be seen in Appendix E, Appendix F, Appendix G, Appendix H, and Appendix J.

Type was always displayed in the color black, although the crisp anti-alias had pixels of gray varying from pure black to the background color which was 1% gray. The untrained human eye cannot detect change in tone of less than 2% (Burke, 1994).

Postscript file fonts were used in the development and delivery of the instrument. As previously discussed, bitmap fonts have no anti-alias capability while TrueType fonts include a variety of optimizations that are likely to hide or otherwise affect the treatments and results of this study.

All content from the instrument was rasterized with Adobe Photoshop on a machine without TrueType fonts and without 8, 10 and 12 point size bitmap fonts. As such, the only fonts available to Photoshop and therefore included in the instrument are Postscript

fonts. Rendering the fonts during development of the instrument, instead of when the test was taken, carefully controls for type format and improves control of type size.

The postscript fonts were taken from Adobe Font Folio version 8. Using file fonts standardized in this way may not be as attractive in some variations but creates a level playing field that allows the purest comparison of the variables in this study.

The first screen of the instrument collected basic demographic information (e.g. a unique participant identifier, year of birth, college major, gender, academic level, and any constraints on vision when the test was taken). All type that was not manipulated as part of the instrument was presented in Dialog with no anti-alias (orthochromatic mode). Dialog is a Postscript typeface from Linotype GmbH. It is a typeface which combines serif and sans-serifs attributes and qualities with a modern look. Dialog can be successfully used with a variety of other faces.

The speed of reading component was an electronic version of the Chapman-Cook Speed of Reading Test. Twenty-four short paragraphs were shown to each participant, one at a time. The order of the 24 items was randomized. As can be seen in Appendix H each question was centered on the left portion of the screen. Centered on the right side were short instructions to identify “which word spoils the meaning of the paragraph.” Beneath these instructions were five possible answers. When possible, the words were chosen from the second half of the paragraph because the spoiled word was always in that part of the paragraph. Participants were instructed to press the number that precedes the answer (1, 2, 3, 4 or 5). The data collection process automatically graded the responses. A participant identifier, type attributes, answer, and the time spent on each item were sent to a local server.

Once participants completed the speed of reading test, they began the comprehension test. Taken from the 1995-96 GRE General Test Descriptive Booklet (Rozmiarek et al., 1996) of verbal comprehension, this electronic variant had students read a short article and then answer four questions about what was inferred or implied. Participants were asked to work quickly and accurately but were allowed to read each article at their own pace. Participants selected and answered each question in order to proceed to the next item. Once they had indicated that they had read the passage (see Appendix L for all the items and Appendix I for a sample passage in a layout), the first question about the passage appeared on the right hand section of the screen (see Appendix J for a sample question and passage). As in the speed of reading section, participants pressed the answer key corresponding to the answers being presented. Once the student answered, the next question was shown. Having answered all four questions relating to a specific passage, the instrument advanced to the next passage. Each participant read four passages and answered four questions about each passage. The order of the four separate segments (a passage and four related questions) was randomized as well as the questions within each segment (the questions associated with each passage). Students were required to answer each question as no mechanism to pass the question was provided. The instrument graded each question as the data was processed. A participant identifier, type style, and answer for each item was reported to a database on the local server.

Completion of the demographic section took most students approximately 3.5 minutes to complete; the speed of reading segment was usually completed in five minutes while the comprehension section took most participants 25 minutes.

Treatment-by-treatment Descriptions

All of the treatments were very similar. The only change between and amongst them was the typographic style. The application randomly determined which typeface and size each participant was subjected to. The choice of typeface and size was consistent for both the speed of reading and comprehension tests.

For the speed of reading test, half of the questions were presented in the "none" style (orthochromatic) and half of the questions were presented with a "crisp" anti-alias. Questions of each rendering technology were grouped so that the participant encountered only one change of style during the speed of reading test. Which style the user received first was determined randomly by the instrument.

Likewise, two comprehension passages and associated questions were presented in Photoshop's "crisp" anti-alias type and two in Photoshop's "none" rendering technology. Groups of passages and their related questions were randomly presented by the instrument.

Results

Introduction

This study aimed to determine what relationships exist, if any, among type size, font selection and type rendering technology with regard to legibility as measured by speed of reading and comprehension when type is read on a computer screen. Two experiments have been performed. One tests the dependent measure of speed of reading (an indicator of legibility) while the other tests reading comprehension.

Data Analysis

Once the data were collected, it was statistically analyzed. Two Repeated Measures Analysis of Variance procedures (one within, two between) were used to determine if significant differences existed (Ary, Jacobs, & Razaveih, 1996; Ott, Larson, & Mendenhall, 1987). Font selection and font size were considered across participants while the rendering technology is considered within participants. The analysis considered the three main effects (selection, size, rendering), first order interactions (selection x size, selection x rendering, size x technology) and the second order interaction (selection x size x rendering) of the two dependent measures (Howell, 1997). These inferential tests were designed to address the eight specific research questions. Where interactions were observed, appropriate post-hoc tests such as pairwise comparisons and searches for simple main effects were conducted to define and explain the nature of the interactions (Greenhouse & Geisser, 1959).

Research Questions

H1) Does the use of serif type vs. sans-serif type result in differences in reading speed (words per minute)?

H2) Does the use of serif type vs. sans-serif type result in differences in reading comprehension?

H3) Does the choice of type size (8 point, 10 point, 12 point sizes) result in differences in reading speed (words per minute)?

H4) Does the choice of type size (8 point, 10 point, 12 point sizes) result in differences in reading comprehension?

H5) Does the choice of type rendering style (orthochromatic vs. anti-alias) result in differences in reading speed (words per minute)?

H6) Does the choice of type rendering style (orthochromatic vs. anti-alias) result in differences in reading comprehension?

H7) Is there an interaction between or among type selection, type size and type rendering style with regard to speed of reading?

H8) Is there an interaction between or among type selection, type size and type rendering style with regard to reading comprehension?

An alpha level of 0.05 was considered the threshold for significance for the purposes of this study.

Descriptive Speed of Reading Findings

Results from the speed of reading test provide non-aggregated mean scores in 12 categories and aggregated scores for 8 additional categories (see Table 6). Type rendering technology is a repeated measure while font selection and type size are randomized across participants. As such, there is no grand mean.

Table 6
Time to Complete Speed of ReadingTasks

Render	Font	Size	*M*	*SD*	*N*
No AA	Helvetica	8	11473	3440	18
No AA	Helvetica	10	9435	3228	18
No AA	Helvetica	12	7497	2483	17
No AA	Palatino	8	12169	5721	21
No AA	Palatino	10	12093	3273	18
No AA	Palatino	12	8962	2057	18
Crisp AA	Helvetica	8	12029	4246	18
Crisp AA	Helvetica	10	11172	3566	18
Crisp AA	Helvetica	12	9721	4046	17
Crisp AA	Palatino	8	10362	3371	21
Crisp AA	Palatino	10	10920	3687	18
Crisp AA	Palatino	12	10162	3200	18
No AA	Helvetica	(all)	9538	3421	53
No AA	Palatino	(all)	11132	4294	57
Crisp AA	Helvetica	(all)	10998	4000	53
Crisp AA	Palatino	(all)	10475	3376	57
No AA	(all)	8	11848	4759	39
No AA	(all)	10	10764	3476	36
No AA	(all)	12	8299	2345	35
Crisp AA	(all)	8	11131	3841	39
Crisp AA	(all)	10	11046	3577	36
Crisp AA	(all)	12	9948	3588	35
No AA	(all)	(all)	10364	3962	110
Crisp AA	(all)	(all)	10727	3682	110

Average time to complete 12 Chapman-Cook test items, times given in ticks (60ths of a second).

The Chapman-Cook test is intended to measure speed and therefore legibility. The items are designed to be cognitively simple and yield very high scores. Table 7 illustrates that the correct response rate was 96.52%.

Table 7
Descriptive Statistics Score for Speed of Reading

Item	*N*	Correct	*M*	*SD*
Chapman A	110	109	.9909	.0949
Chapman B	110	106	.9636	.1872
Chapman C	110	108	.9818	.1336
Chapman D	110	106	.9636	.1872
Chapman E	110	109	.9909	.0949
Chapman F	110	101	.9182	.2741
Chapman G	110	108	.9818	.1336
Chapman H	110	108	.9818	.1336
Chapman I	110	106	.9636	.1872
Chapman J	110	105	.9545	.2083
Chapman K	110	102	.9273	.2597
Chapman L	110	98	.8909	.3118
Chapman M	110	106	.9636	.1872
Chapman N	110	107	.9727	.1629
Chapman O	110	109	.9909	.0949
Chapman P	110	108	.9818	.1336
Chapman Q	110	106	.9636	.1872
Chapman R	110	105	.9545	.2083
Chapman S	110	106	.9636	.1872
Chapman T	110	106	.9636	.1872
Chapman U	110	107	.9727	.1629
Chapman V	110	107	.9727	.1629
Chapman W	110	108	.9818	.1336
Chapman X	110	107	.9727	.1629
Average	110	106.17	.9652	.1834

Comprehension scores on the Chapman-Cook speed of reading are very high, as expected, because the test is meant to be cognitively simple. The means listed above indicate the fraction of students who correctly answered each item.
N = 110. 2640 total responses.

Inferential Speed of Reading Findings

The repeated measure comparison of means reveals a statistically significant main effect with regard to size and an interaction between font selection and type rendering technology with regard to speed of reading.

Type size.

Type size, as reported in Table 8, was found to be a significant main effect. ($F(2, 104) = 6.582$, $p = 0.002$). This result leads to the rejection of null hypothesis H3 and forces a conclusion that size does matter.

Note that eta squared is 0.112 and the statistical power is 0.903. In this case we have some impact of effect and extremely high power.

A least significant difference pairwise post-hoc analysis, see Table 9, indicates that 12 point type is read faster than either 8 or 10 point type onscreen ($p < 0.05$). However, the difference between the reading speed of 8 point and 10 point type is not statistically significant ($p = 0.375$). Of the three type sizes investigated, 12 point type is read fastest.

Table 8
Analysis of Variance for Speed of Reading

Source	*df*	*F*	η^2	Power	*p*
		Between Subjects			
Size	2	6.582	.112	.903	.002**
Font	1	.935	.009	.160	.336
Size x Font	2	.894	.017	.201	.412
error	104				
		Within Subjects			
Render	1	1.074	.010	.177	.302
Render x Size	2	2.462	.045	.485	.090
Render x Font	1	5.934	.054	.675	.017*
Render x Size x Font	2	.474	.009	.126	.624
error	104				

* p < .05 level (2-tailed). ** p < .01 level (2-tailed).

Table 9
Pairwise Comparison for Speed of Reading with regard to Size

(I) Size	(J) Size	Mean Difference	*SE*	*p*	95% Confide Lower Bound
8 points	10 points	603.437	676.855	.375	-738.793
8 points	12 points	2397.586	681.997	.001**	1045.160
10 points	12 points	1794.150	694.325	.011*	417.276

* p < .05 level (2-tailed). ** p < .01 level (2-tailed).

Font selection.

No main effect was observed for font selection, see Table 8 again, with regard to speed of reading ($F(1, 104) = 0.935$, $p = 0.336$). Therefore, null hypothesis H1 cannot be rejected. Neither Helvetica nor Palatino shows a clear advantage of reading speed when other variables are not taken into effect.

The effect size is very small at 0.009 and the power is also small at 0.160.

Type rendering technology.

No main effect was observed for type rendering technology with regard to speed of reading ($F(1, 104) = 1.074$, $p = 0.302$), as can be seen in Table 8. Null hypothesis H5 cannot be rejected. Speed of reading is not significantly better for either rendering technology when interactions are not taken into account.

The eta squared indicator of effect size is 0.010 and the statistical power is 0.177. There seems to be little effect to find and little power to find it with.

Mauchly's Test of Sphericity was computed. Based on the value of that test (Mauchly's $W = 1.000$, approximate Chi-Square = 0.000), the Greenhouse-Geisser test values were not distinct from the values when sphericity was assumed.

Interaction between font selection and type rendering technology.

Type rendering technology and font selection interact ($F(1, 104) = 5.934$, $p = 0.017$). As such, post-hoc analysis is needed to further explain the data.

The first step in this analysis is to collapse the data set to consider only the two variables in the interaction. A cursory look at the means in Table 10 suggests that the means for Helvetica type is read faster when rendered in orthochromatic mode while Palatino appears to be read faster in anti-alias mode. Although Helvetica appears to be read faster than Palatino in orthochromatic mode, in anti-alias mode the reading speeds seem very similar. However, we can not base our analysis on descriptive statistics.

Table 10
Post-Hoc Descriptives for Speed of Reading Interaction between Font Selection and Rendering Technology

Interaction Style	*N*	*M*	*SD*	*SE*	95% Confider Lower Bound
Helvetica, None	53	9537.6226	3421.2284	469.9419	8594.6153
Helvetica, AA	53	10997.6981	4000.1919	549.4686	9895.1087
Palatino, None	57	11132.4737	4294.3403	568.7993	9993.0318
Palatino, AA	57	10475.0702	3375.6738	447.1189	9579.3833

It is necessary to compare more than descriptive means to determine which table cells actually indicate statistically significant differences. One way to accomplish this is through a comparison of simple effects (Keppel & Zedeck, 1989). By performing a simple test of each row and column of the interaction matrix, it can be determined where the effects of the interaction occur. Essentially, an analysis of font selection is conducted at both the anti-alias level and none level. An analysis of rendering technology is conducted at both the Helvetica level and Palatino level. Because of the nature of this analysis, the repeated measure must be broken. This causes each participant to appear in the new data set twice (once for the anti-alias and once for the none rendering style).

Table 11
Simple Effects for Speed of Reading Test Interation between Font Selection and Rendering Technology

		df	*F*	η	η^2	*p*
Simple Effect of Font Selection at Anti-Alias Rendering	Between	1	0.551	.071	.005	.460
	Within	108				
Simple Effect of Font Selection None Rendering	Between	1	4.596	.202	.041	.034*
	Within	108				
Simple Effect of Rendering Technology at Helvetica	Between	1	4.078	.194	.038	.046*
	Within	104				
Simple Effect of Rendering Technology at Palatino	Between	1	.826	0.086	.007	.365
	Within	112				

* $p < .05$ level (2-tailed).

Table 11 indicates the results of the analysis for simple effects. These tests indicate that Helvetica none is read faster than Helvetica crisp ($F(1, 104) = 4.078$, $p = 0.046$). Similarly, Helvetica none is read faster than Palatino none ($F(1, 108) = 4.596$, $p = 0.034$). The other comparisons were not statistically significant at the 0.05 level.

By comparing means we conclude that Helvetica is read faster in the none (orthochromatic) rendering technology than in crisp (anti-alias). Through the same type of analysis, it is apparent that Helvetica none is read significantly faster than Palatino none. Helvetica none is not significantly better than Palatino crisp. As such, the difference in means may be due to chance in all but two comparisons.

Another technique to determine which of the individual comparisons are significant is to compare subsets with a one-way ANOVA (Cross, 2000; Howell, 1997). This analysis confirms the simple effect findings and also provides two more non-significant comparisons as shown in Table 12.

Table 12
Interaction Post-Hoc Significance
for Speed of Reading

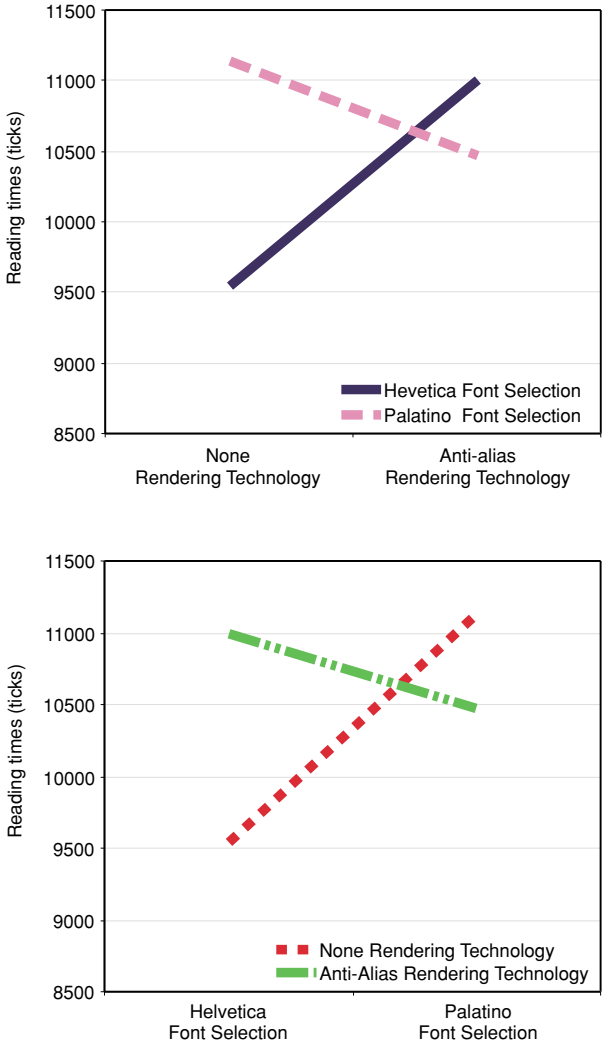
	Helvetica None	Helvetica Anti-alias	Palatino None	Palatino Anti-alias
Helvetica, None	—	.049*	.029*	.197
Helvetica, Anti-alias		—	.853	.471
Palatino, None			—	.356
Palatino, Anti-alias				—

* $p < .05$ level (2-tailed).

The interaction can be seen visually in Figure 14. These graphs illustrate that the reading speed of Helvetica none is dramatically faster than Helvetica crisp and also that Helvetica none is read faster than Palatino none. Although the graphs make it appear as

though Palatino crisp is read faster than Palatino none, this difference in means is not statistically significant ($F(1, 112) = 0.826$, $p = 0.365$).

Figure 14
Speed of Reading Interaction of Font Selection and Rendering Technology, two views



Helvetica none is read faster than Helvetica with a crisp anti-alias and Palatino none although Helvetica none is not significantly faster than Palatino with a crisp anti-alias.

These analyses indicate that there is a significant interaction of font selection and rendering technology with regard to speed of reading. Further, the post-hoc analyses

pinpoint the exact nature of the interaction. As such, we can reject the null hyphothesis H7 and conclude that certain fonts perform better in one rendering style than another.

Descriptive Reading Comprehension Findings

In addition to the speed of reading data, reading comprehension data were collected. Table 13 shows the means by typographic style while Table 14 show the means by each test item .

Table 13
Descriptive Statistics for Comprehension by Typographic Style

Render	Font	Size	*M*	*SD*	*N*
No AA	Helvetica	8	1.33	1.14	18
No AA	Helvetica	10	2.44	1.69	18
No AA	Helvetica	12	2.24	1.39	17
No AA	Palatino	8	1.86	1.11	21
No AA	Palatino	10	2.00	1.03	18
No AA	Palatino	12	2.33	1.19	18
Crisp AA	Helvetica	8	1.61	1.33	18
Crisp AA	Helvetica	10	2.11	1.68	18
Crisp AA	Helvetica	12	2.47	1.50	17
Crisp AA	Palatino	8	2.05	1.63	21
Crisp AA	Palatino	10	2.22	1.40	18
Crisp AA	Palatino	12	1.89	1.75	18
No AA	Helvetica	(all)	2.00	1.48	53
No AA	Palatino	(all)	2.05	1.11	57
Crisp AA	Helvetica	(all)	2.06	1.52	53
Crisp AA	Palatino	(all)	2.05	1.57	57
No AA	(all)	8	1.62	1.14	39
No AA	(all)	10	2.22	1.40	36
No AA	(all)	12	2.29	1.27	35
Crisp AA	(all)	8	1.85	1.50	39
Crisp AA	(all)	10	2.17	1.52	36
Crisp AA	(all)	12	2.17	1.64	35
No AA	(all)	(all)	2.03	1.30	110
Crisp AA	(all)	(all)	2.05	1.54	110

The mean values indicate the average number of items answered correctly by typographic style (rendering technology, font selection and size).

Minimum possible score is 0.
Maximum possible score is 8 for each half of the test (No anti-alias and Crisp anti-alias).

Table 14
Descriptive Statistics for Comprehension by Test Item

	M	*SD*
Comprehension B	0.25	0.44
Comprehension C	0.33	0.47
Comprehension D	0.29	0.46
Comprehension E	0.17	0.38
Comprehension G	0.33	0.47
Comprehension H	0.27	0.45
Comprehension I	0.26	0.44
Comprehension J	0.32	0.47
Comprehension L	0.25	0.44
Comprehension M	0.18	0.39
Comprehension N	0.25	0.44
Comprehension O	0.32	0.47
Comprehension Q	0.25	0.44
Comprehension R	0.25	0.43
Comprehension S	0.35	0.48
Comprehension T	0.26	0.44
Aggregate	4.082	2.099

The mean values suggest what fraction of participants were able to answer each question correctly. For example, 32% of the participants answered question J in the comprehension section correctly.

Items A, F, K and P are not shown in the table because they are passages and not questions.

Minimum score is 0; Minimum possible score is 0. Maximum score is 10; Maximum possible score is 16.

N = 110.

Performance was not as high as expected and the range of scores was less variable than anticipated. Because standardized tests seek to accurately rank participants in terms of achievement, the instruments are designed to generate a wide range of scores, centered on a middle value, with a sufficient degree of variance to rank participants in relationship to each other. It was hoped that the mean score would be 0.50 (indicating that roughly half of the partipants correctly answered each question). However, the means are roughly

half that value (0.26). As such, the range of the data is restricted. This condition makes it more difficult to find actual differences in means.

Reading Comprehension Reliability

Table 15 shows the reliability matrix for each test item. The reliability matrix indicates how well each question correlates with overall student achievement. A high, positive value (near to 1.000) indicates that success on this item is a good predictor of success on the test as a whole. A small value (near to 0.000) indicates that this item is not a good predictor of overall success on the instrument. This is usually an indicator that the item itself is suspect. It is also possible to have a high, negative value which indicates that the questions misleads many students or that the answer key is incorrectly coded. The associated p values indicate whether the correlations for each question are significant or may simply be the result of chance. In general, the reliability indicators are modest. Note that 13 of the 16 items are statistically significant despite the restriction of range problem.

Table 15
Comprehension Instrument Reliability Matrix, Correlations between Items and Aggregated Score

	r	*p*
Comprehension B	.457	.000**
Comprehension C	.334	.000**
Comprehension D	.224	.019*
Comprehension E	.178	.063
Comprehension G	.520	.000**
Comprehension H	.445	.000**
Comprehension I	.046	.636
Comprehension J	.310	.001**
Comprehension L	.307	.001**
Comprehension M	.309	.001**
Comprehension N	.407	.000**
Comprehension O	.235	.014*
Comprehension Q	.307	.001**
Comprehension R	.129	.178
Comprehension S	.484	.000**
Comprehension T	.233	.014*

* p < .05 level (2-tailed). ** p < .01 level (2-tailed).

N = 110.

Inferential Comprehension Findings

The repeated measure comparison of means did not reveal main effects for type size, font selection or rendering technology, first level interactions (size * font, size * render, font * render) or second level interactions (size * font * render) with regard to comprehension. As such, none of the null hypotheses can be rejected.

Type size.

As seen in Table 16, no main effect for type size was observed for type size with regard to comprehension ($F(2, 104) = 2.950$, $p = 0.057$). As such, null hypothesis H4 cannot be rejected. A comparison of means of versions of the instrument that used different type sizes did not result in enough difference to discount chance. A strict

interpretation of the inferential statistics suggests that type size is not a factor in the comprehension of type.

Table 16
Analysis of Variance for Comprehension

Source	*df*	*F*	η^2	Power	*p*
		Between Subjects			
Size	2	2.950	.054	.563	.057
Font	1	0.015	.000	.052	.904
Size x Font	2	1.382	.026	.291	.256
error	104				
		Within Subjects			
Render	1	0.018	.000	.052	.895
Render x Size	2	0.330	.006	.102	.719
Render x Font	1	0.036	.000	.054	.850
Render x Size x Font	2	0.891	.017	.200	.413
error	104				

No significant findings were observed with regard to comprehension.

While the particulars of this experiment did not obtain statistical significance ($p = 0.057$), the effect size as indicated by partial eta squared was 0.054 while the power indicated was 0.563. In this case, even a modest improvement in statistical power could result in a p value less than 0.05. Factors such as increasing the number of participants could produce differing results. Cohen (1988) reminds us that the widely used 0.05 level of significance is an arbitrary convention.

An overly conservative analysis may create Type I errors. This error is no less egregious than Type II errors (Howell, 1997).

Practitioners must choose a type size and there is no penalty for making a choice with only weak statistical support. For this audience, some would argue that an alpha

greater than 0.05 is appropriate (P. E. Doolittle, personal communication, September 21, 2001).

Font selection.

Table 16 also shows that no main effect is observable for font selection with regard to comprehension ($F (1, 104) = 0.015$, $p = 0.904$). As such, null hypothesis H2 cannot be rejected. Choice of font (Helvetica versus Palatino) does not affect comprehension.

The observed effect size for font selection with regard to comprehension is negligible (<0.001) while the power is weak (0.052).

Type rendering technology.

Likewise, no main effect was observed for font selection with regard to comprehension ($F (1, 104) = 0.018$, $p = 0.895$). The within subjects portion of Table 16 illustrates the lack of significance. As such, null hypothesis H6 cannot be rejected. Rendering technology did not affect achievement in this study.

The effect size for rendering technology as indicated by eta squared is negligible at less than 0.001 while the power is weak at 0.052.

Mauchly's Test of Sphericity was computed. Based on the value of that test (Mauchly's $W = 1.000$, approximate Chi-Square = 0.000), the Greenhouse-Geisser test values were not distinct from the values when sphericity was assumed.

Interaction with regard to Comprehension.

There are four possible interactions with regard to comprehension: type size x font selection, type size x rendering technology, font selection x rendering technology, and type size x font selection x rendering technology. None of these four possible interactions, however, is statistically significant (see Figure 16). As such, we can not reject null hypothesis H8. No interaction between the three independent measures was observed for comprehension.

Analysis Summary

The results indicate that for legibility, as measured by speed of reading, larger type is often read faster than smaller type. Also, an interaction was found between font selection and rendering technology. Helvetica is read faster when rendered without an anti-alias than Helvetica with an anti-alias; Helvetica is also read faster than Palatino without an anti-alias. With regard to reading comprehension, no significant findings were found and the null hypotheses cannot be rejected.

Discussion

Background

This study seeks to determine how font selection, type size and type rendering technology affect the legibility of onscreen type and comprehension. Two instruments were developed: one that tests for speed of reading, widely regarded as a good indicator of legibility; and another that tests verbal comprehension. By administering these instruments to 117 participants, data were collected on the relationship between the factors of interest.

Statistical analysis determined that there was a main effect for type size with respect to speed of reading and therefore legibility. The larger type in the study, 12 points, was read faster than either 10 point type or 8 point type although there was no significant difference when comparing 8 point type to 10 point type individually.

Analysis also revealed that there was a statistically significant interaction between font selection and type rendering technology. Helvetica was read faster when rendered without an anti-alias than with an anti-alias. Helvetica without an anti-alias was read faster than Palatino without an anti-alias. However, Helvetica without an anti-alias was not read statistically faster than Palatino with an anti-alias. Nor was there a significant difference between the reading speed between Palatino with an anti-alias and Palatino without an anti-alias.

When considering verbal comprehension, no statistically significant main effects or interactions were discovered. This confirms a widespread belief that legibility indicators

such as speed of reading are more likely to find small differences in type presentation than comprehension indicators (Manzel & Tranel, 1995; Tinker, 1963).

Discussion of Speed of Reading Results

Type Size

Type size was found to play a role in speed of reading. The type rendered at 12 point size was read faster than either 8 or 10 point type onscreen ($p < 0.05$). However, there was no statistical difference between the 8 point size and 10 point size ($p = 0.375$).

Geske's (2000) study of onscreen type size used a range of type in sizes larger than this study; however, Geske also reported that 12 point was the fastest read size. However, we should be cautious in comparisons of point size within and between studies. The numeric identification of type is based on historical factors and is not a good indicator of the perceptual size. Using the number of pixels in the majority of upper case letters (the cap height) or the lowercase letters (the difference between the baseline and the x-height) gives a better indication of perceptual size. As such, studies should scale their type to a standard size (this study used Helvetica as the standard) and be cautious about extrapolations that have little or no meeting.

This study's finding that of fonts with equal pixel sizes, 12 point type is read faster than 8 or 10 point type is not surprising. It is logical to assume that moderately larger typefaces will be read faster because of the availability of a larger matrix of pixels for each letter to be defined and also because each glyph uses a larger degree of the visual arc. There are a predetermined number of rods and cones in the human eye to interpret visual symbols such as type glyphs. Moderately larger type, such as that used in this

experiment, will result in more optical receptors being used for reading an onscreen glyph. It is also possible that more receptors are needed for the pattern matching of onscreen type because of the unusual attributes of this presentation method. Factors such as additive versus subjective light (projected light versus reflected light), low resolution, refresh rate and difficulty in manipulating angle of reading may also impact reading onscreen.

Practitioners should consider that this and other studies have found that onscreen type is read fastest at 12 point size. Although screen real estate is a limited and valuable commodity, larger text is read faster and is therefore more legible. Although developers of electronic content are probably unwilling to increase the size of functional, navigation elements at the cost of reducing the amount of content in a page—they should at least consider formatting content in ways that are most conducive to legibility. Even if type used for navigation remains small (and read somewhat more slowly), it would be of great assistance to format long galleys of text in an appropriate point size (roughly equal to 12 point Helvetica or Palatino) to facilitate ease and speed of reading. Practitioners should consider that Geske (2000) found that speed of reading increases (longer time periods) for type larger than 12 points. In the absence of strict controls on the relative size of type, initial results suggest 12 points is the optimal type size for reading some fonts onscreen.

Font Selection

No main effect was found with regard to font selection. In fact, there is little existing research with regard to onscreen font selection and speed of reading. Geske (2000) and Weisenmiller (1999) found no significant differences for speed of reading with regard to

font selection. Grant found a large, meaningful and statistically significant difference in means: serif faces were read faster in his study (2000).

It is interesting that Grant's results differ markedly from this study and others in the literature. However, Grant appeared to offer only the most basic of controls: although he specified preferred fonts for use by the web browser, that study did not render the type during development or control for font format (bitmap versus TrueType versus Postscript). Grant's study may be affected by some uncontrolled variable and therefore not be clear evidence of serif faces being superior to faces without serifs. Although Grant's study is a good test of real-world reading conditions, it does not offer precise control of secondary variables.

With some evidence that serifs are read faster onscreen and other evidence that there is no appreciable difference, many practitioners would likely conclude that serif type has a preponderance of the evidence. However, that conclusion differs from almost all expert opinion. Further, there is an interaction (see Interaction between Font Selection and Type Rendering Technology) which makes a contradictory claim.

Type Rendering Technology

Type rendering technology did not indicate a main effect with regard to legibility in this study.

As discussed extensively in the review of literature, both typographic and computer experts are widely divided on the value of anti-alias rendering when compared to orthochromatic rendering.

Weisenmiller's (1999) study did not expressly set out to study rendering technology; however, his choice of software offered a compelling comparison of Acrobat's anti-alias, orthochromatic rendering on screen and 600 dpi laser printer output. It is worthy of note that Acrobat's anti-alias is different than the crisp anti-alias used in this study and the orthochromatic rendering in Weisenmiller's study is different from the orthochromatic rendering used in this study. Weisenmiller found that his anti-alias type was read onscreen faster than his orthochromatic type regardless of typefaces. Many of Weisenmiller's typefaces were deliberately optimized for use onscreen and none of them were likely available only in Postscript versions as in this study.

One conclusion is that type optimized for use onscreen performs equally well (or near equally well) when rendered with an anti-alias while type not optimized for use onscreen does not read faster when anti-aliased. This would explain why Weisenmiller's type (most likely TrueType) showed a benefit in anti-alias mode that this study's type did not (Postscript). It is also possible that the differing anti-alias technologies (Acrobat anti-alias versus Photoshop's crisp) played a role.

The benefits and costs of anti-alias technology are still not known. However, the significant interaction between factors does offer a more complex view of the rendering technology and its impact.

Interaction between Font Selection and Type Rendering Technology

An interaction was found for speed of reading with regard to font selection and rendering style. Helvetica none (Helvetica rendered in orthochromatic mode with no anti-alias) was found to be read faster than Helvetica crisp (anti-alias) and faster than Palatino

none. The other possible comparisons for simple effects revealed no significant differences.

These findings suggest that the use of anti-alias technology is a moderating factor in speed of reading. While the reading of simple letter glyphs of Helvetica (a sans-serif typeface) are slowed by the crisp anti-alias, the tonal resolution offered by anti-alias does not hinder the display of the more delicate serifs and letterforms of Palatino ($p = 0.356$). Perhaps most surprising, Helvetica none is not read faster than Palatino anti-alias ($p = 0.197$). This lends some support for the idea that an anti-alias rendering style acts to improve tonal resolution and can help complex typefaces be rendered at low resolutions. The crisp anti-alias rendering technology should be considered contextually as it can slow reading speed; however, it may also have no impact or improve reading speed.

Weisenmiller (1999) found that anti-alias performed better as a main effect while the current study found that performance was affected by font selection and rendering technology. The current study found that Helvetica performs best without an anti-alias. To the contrary, there is no significant difference in reading speed when Palatino is used.

None of Weisenmiller's results suggested that Helvetica would have been read faster without an anti-alias. Additionally, Weisenmiller's results predicted that Palatino would be read faster when rendered with an anti-alias. This study found that there was no significant difference between the two renderings of Palatino; nonetheless, Helvetica none was read faster than Palatino none and was not read significantly faster than Palatino anti-alias which indicates there may be some meaningful difference between Palatino across the rendering styles. This is evidence that the non-significant differences when comparing the Palatino data is likely to be a Type II error.

It is also worthy of note that Grant (2000) reported that serif type was read faster than sans-serif type. His findings were not supported by this study which found no significant differences between the best rendering technology for the serif face (Palatino crisp) and the best rendering technology for the sans-serif face (Helvetica none). As previously discussed, this could be attributable to differences in file formats. Another explanation could be due to differences in the typefaces chosen to represent the two families. Many of Grant's participants viewed Ariel instead of Helvetica and most viewed Times New Roman or Times instead of Palatino.

Because of disparities between Weisenmiller's findings and those of the current study, there are no clear recommendations for developers of interactive content. According to Weisenmiller's study, all onscreen type should be rendered in anti-alias mode. This is in stark contrast to this study, which suggests that sans-serif fonts (like Helvetica) should be rendered without an anti-alias while the difference between rendering serif fonts (like Palatino) is not signficant. A precise comparison of the two studies would suggest that when TrueType fonts will be used in Adobe Acrobat, sans-serif faces should be anti-aliased; and when Postscript fonts will be used outside of Adobe Acrobat (Acrobat only offers anti-alias rendering), sans-serif faces should not be anti-aliased.

The context of these suggestions is also important. There is no evidence to suggest, and much to the contrary, that type created in an appropriate font and in an appropriate size cannot be read whether or not it is anti-aliased. The difference between the most legible style (Helvetica none) and the least legible style (Palatino none) is a 16% increase in reading time (computed by taking the time to read Palatino none divided by Helvetica

none). In fact, if there is any evidence to suggest that one style of type is more legible than another, practitioners should take this evidence under advisement.

Discussion of Comprehension Results

Restriction of Range

The mean performance on the verbal comprehension items was 4.082 out of a possible 16.000. The overall standard deviation was 2.099. This average was lower than expected. As such, the data may suffer from a restriction of range.

There are several factors that may explain the lower than anticipated verbal comprehension achievement. First, the students were recruited as a convenient population of student volunteers. The majority of students in the data set were recruited from sophomore-level classes (85 students from sophomore classes, 3 students from junior classes, and 22 from senior classes). Students enrolled in lower-level classes may not be as academically prepared as the typical Graduate Record Exam (GRE) test-taker. Second, because the GRE is used primarily to determine which students will be accepted to graduate school, it is possible that only the best students take the GRE at all. In effect, the students who typically take the GRE self select. Third, students taking the GRE are likely to be highly motivated. This same level of motivation may not be present in student volunteers with no stake in their performance.

Type Size

This study did not find a main effect or interaction involving type size (8 points, 10 points and 12 points) with regard to comprehension. This finding is in conflict with

Geske's study, which found that 12 point type resulted in greater recall than 14 point or 10 point (2000).

In the absence of evidence to contradict Geske's recommendations, it is prudent to continue to set type in 12 point size. Further, the current study did determine that 12 point type is read more quickly and is therefore more legible. This further bolsters the claim that 12 point is the optimal size for onscreen type.

Font Selection

Font selection was not found to provide a main effect with regard to comprehension. This finding confirms prior research by Geske (2000) and Weisenmiller (1999).

In the absense of significant results with regard to font selection and comprehension, practitioners are free to use their own judgement. Additionally, practitioners should carefully consider the impact of speed of reading and the roll that font selection plays in those typographic decisions (see Font Selection and Interaction between Font Selection and Type Rendering Technology in the Discussion of Speed of Reading Results section).

Type Rendering Technology

This study supports Weisenmiller's (1999) findings that rendering style was not a factor in reading comprehension. Although the font file formats and fonts used were different, neither study found rendering technology to be a factor in comprehension.

Again, in the absence of decisive findings with regard to comprehension, developers of onscreen content are left with only the legibility findings for guidance (see Type

Rendering Technology and Interaction between Font Selection and Type Rendering Technology in the Discussion of Speed of Reading Results section).

Discussion At Large

This study contradicts many findings of prior studies. What is not known continues to greatly outweigh what is known about type onscreen. When compared to the massive bodies of analysis regarding type in print by Miles Tinker and his contemporaries, it is easy to ascertain how recent and underdeveloped research into onscreen type actually is.

Much of the variability in this and other studies might be attributable to a wide range of variables. Computer applications have variances in hardware resolutions, software resolutions, typefaces, file formats for typefaces, type sizes and rendering technologies. For example, this study compared a crisp anti-alias style to a none anti-alias style (orthochromatic mode). While there is reason to believe that these styles are very similar to many other anti-alias styles, insufficient research has been done to confirm that this is the case. The differences between Weisenmiller's study, which found Acrobat's anti-alias improved reading speed of sans-serif and serif fonts, and this study, which found anti-alias hampered the sans-serif font, might be attributable to differences in the anti-alias itself. This particular factor is exacerbated by the fact that emerging anti-alias technologies seem to be radically different in appearance than earlier anti-alias rendering technologies. Likewise, the difference in font shapes and styles may make comparison and aggregation of research more difficult. This study used Palatino as the serif face while two earlier studies used Times and found very different results. Perhaps the serif and sans-serif families are not meaningful and equivalent groupings onscreen. Clearly

these problems will only be elucidated with further research carefully designed to replicate and build on prior research.

This study offers some evidence that legibility as indicated by speed of reading is a more delicate indicator of typeface than comprehension. This confirms a wide body of research in print (Manzel & Tranel, 1995; Tinker, 1963) and also confirms Weisenmiller's work. However, Geske (2000) was surprised to find that comprehension indicated more significant results than speed of reading.

Clearly, this study offers support for some previous findings, addresses areas that have previously been under-researched and raises some new questions. A better understanding of these questions will come through continued research.

Areas of Future Research

A plethora of areas exist for future research regarding onscreen type. The field is still immature with only a few experiments revealing often-conflicting findings. When compared to the massive body of print work undertaken in the first 50 years of the 20th century, it is obvious that more research is needed. The well-reasoned body of work into type in print can also provide insight into what types of research should be undertaken.

Future research should replicate. A single study offers only a glimpse into the behavior of onscreen typography. Studies of onscreen typography should be replicated.

One area of expansion could be to consider more typefaces. While this study used two popular, available and representative faces—thousands of other typefaces remain unstudied. In fact, it is not definitive whether or not serif and sans-serif fonts behave onscreen as two distinct styles of type.

Even for two similarly designed and named fonts, it is not known whether the file formats plays a role in the legibility and comprehension of type. It is possible that the hinting available in TrueType versions of type will interact with render technology, font selection and size. Emerging file formats like the open type format (OTF) may also behave differently.

The anti-alias format itself is another factor. Although there is evidence to believe that many of the current anti-alias styles are similar in overall appearance and tonal quality, newer technologies are increasingly available. New styles of anti-alias rendering are becoming available on handheld devices. Another example is the perceptual encoding anti-alias available in Apple's Mac OS X. Not only does this technology apply more computing power to the computation of the anti-alias but it also modifies the letterforms themselves to more precisely match the orthogonal matrix of the raster screen. Although it now appears that anti-alias rendering will be used in the majority of all text in the future, it is still important to know the ramifications and consequences of different rendering choices.

In this study, there were no significant findings with regard to comprehension. Other studies, however, have found comprehension findings significant even when no differences in legibility were found (Geske, 2000). The use of similar instruments, which separate legibility from comprehension, should be encouraged.

Resolution is another avenue of study. Logical screen sizes are increasing over time. That is to say that the number of pixels available on an average screen increases over time. The impact of the large real estate is not known. Similarly, the concentration of these pixels varies (physical size). For example, Apple's 15" Studio Display offers the

same 1024x768 pixels as the 12" iBook dual-port display. One might expect that the same text, rendered the same way might be read differently. This factor is particularly compelling as different software and platforms deal with dot pitch in radically different ways.

Research has also neglected different monitor technologies. None of the papers consulted in this study make mention of LCD (liquid crystal displays). They are not only the standard for laptop computers, but they are also making serious inroads into the desktop market as well.

There is also a digital renaissance with regard to bitmap font file formats. Although TrueType, Postscript and OTF fonts have largely replaced bitmaps—many typographers continue to develop them. One area in which bitmap fonts are again being widely used is on the world wide web where some designers feel that only a hand-tweaked bitmap face can communicate clearly and at small size. The impact of bitmap fonts is not known although this study's findings of fast reading speed with orthochromatic Helvetica text provides some support for the idea of using bitmap fonts effectively.

Clearly, there is room for substantial research in the area of onscreen typography. If our society continues to migrate towards onscreen reading, we will need to know more about the process.

Future Research should Exercise Appropriate Controls

This study reviewed the limited literature regarding onscreen typography. Previous studies, however, failed to either properly document controls for extraneous variables or

failed to control for extraneous variables. Having talked to many of the authors, the latter seems more common than the former.

Future research should deliberately document and control for variables not independently manipulated. In fact, the contradictory findings of research thus far may be attributable to the lack of tight controls. Studies should consider, control and document in a number of key areas:

- monitor type (CRT vs. LCD)
- monitor size (viewable size and aspect ratio)
- monitor resolution (ppi and dot pitch)
- operating system resolution
- rendering style (fractional width setting; orthochromatic, 3x anti-alias type, 4x anti-alias type, or perceptual encoded anti-alias)
- font technology (bitmap, TrueType w/ hinting, Postscript, OpenType)
- type sizes (comparable font to font, consider x-height and cap height not point size)
- typefaces (which faces are used, manufacturer)
- reading angle and attack
- character based attributes (bold, italic, track, lead, ...)
- viewing distance
- viewing angle
- visual acuity of participants (normal, corrected, uncorrected)
- figure/ground (color of type and background)
- ambient lighting

When extraneous variables are controlled, the research will be more sound and the generalizations made regarding that research will also be more sound. It is also possible that tighter controls will allow for more consistent results between studies. By comparing the results of many carefully controlled studies, common themes are likely to emerge. Over time intersections between and among variables will begin to provide a reasonable framework to help researchers and practitioners understand what combination of attributes result in most legible onscreen type and what factors, if any, result in the greatest comprehension.

Implications for Practice

One should be exceedingly cautious in making recommendations for practice. The participants of this study represent a tiny subset of students at a single university in the Blue Ridge mountains. Further, the students are largely from a single major. It would be unwise to suggest that these results prove anything about reading onscreen for a college population at large, much less most computer users.

This study has shown that small typefaces are not read as quickly as larger ones. And while reading speed is by no means the only consideration of providers of computerized content and instruction, it does lend evidence to the argument made by Geske that there may be optimal typeface sizes.

The interaction between font selection and type rendering style is even more muddy. There is evidence to suggest that the use of rendering style is highly contextual and should be considered on a case by case basis, at least until further research provides

greater clarity to the conditions which make each style optimal. Perhaps one relevant finding is that it appears as though the use of either rendering technology explored here does not have a meaningful impact on comprehension and does not dramatically impair reading even when legibility is impacted. The degradation of the non-ideal rendering style seems no more dramatic than the choice of a less-optimal typeface in print.

Content creators are left with several suggestions (see Discussion of Speed of Reading Results) as well as the advice to carefully consider and evaluate their online typographic choices. Creators should realize that new rendering technologies are not uniformly positive or negative. Further, creators are encouraged to conduct usability studies to test typeface styles in their particular environment.

Summary

This study considered the relationships between type size, font selection, and type rendering technology of legibility as indicated by speed of reading and comprehension when testing was performed on a computer with a CRT. The results offer an additional view on a new, under-researched area. The finding that there are optimal sizes onscreen will be an important concept to some. Evidence that the effect of anti-alias is highly contextualized based on type family and potentially specific typefaces could have an even more far-reaching impact. Clearly the role of reading on the computer needs to be better understood and explored through future research.

References

Adobe. (2000a). Acrobat Reader (Version 4.0.5). San Jose, CA: Adobe Systems.

Adobe. (2000b). Adobe Photoshop (Version 6.0) [Raster Image Editor]. San Jose, CA: Adobe Systems.

Adobe. (2000c). Adobe Type Manager (Version 4.6) [Macintosh Control Panel]. San Jose, CA: Adobe Systems.

Apple. (1991). *TIL 6492: Same-Size Images Not Possible with Different Resolutions*. Retrieved November 25, 2000, from http://til.info.apple.com/techinfo.nsf/artnum/n6492

Apple. (1996). Mac OS (Version 8.5) [Operating System]. Cupertino, CA: Apple Computer.

Apple. (1998). *TIL 56386: Bitmaps: Basic Editing Concepts*. Retrieved November 25, 2000, from http://til.info.apple.com/techinfo.nsf/artnum/n56386

Apple. (2000a). *Apple Studio Display 17"/16" vis (2000)*. Retrieved November 26, 2000, 2000, from http://store.apple.com/1-800-MY-APPLE/WebObjects/AppleStore.woa/184/wo/0c4ce1FsAOgs7WO9B4/2.5.0.3.28.1.13.0.13.0

Apple. (2000b). *The graphics industry enters the Quartz age*. Retrieved November 28, 2000, from http://www.apple.com/macosx/technologies/graphics.html

Apple. (2000c). *TIL: Multimedia Glossary RS-Z, N16537*. Retrieved November 24, 2000, from http://til.info.apple.com/techinfo.nsf/artnum/n16537

Armstrong, J., Herbert, B., & Gowin, S. (2000). *MacroMedia Director: 8 using Director*. San Francisco, CA: MacroMedia.

Ary, D., Jacobs, L. C., & Razaveih, A. (1996). *Introduction to research in education* (5th ed.). Orlando, FL: Harcourt Brace College Publishers.

Bayley, P. (2000). *OS X: A merging of various decades-old technologies. Does it work?* Retrieved September 21, 2000, from http://www.xappeal.org/archive/osxmerge.shtml

Bradshaw, A. (2000). *Maximizing visual learning*. Ames, IA: International Visual Literacy Association).

Bringhurst, R. (1999). *The elements of typographic style* (2nd ed.). Vancouver, B. C., Canada: Hartley and Marks.

Buckingham, B. R. (1931). New data on the typography of textbooks. In *Yearbook of the National Society for the Study of Education* (Vol. 30, pp. 93-125).

Burke, J. (1994). Connections 2 [VHS videocassette]. Bethesda, MD: The Learning Channel.

Canary, M. R. (1983). *An analysis of typographic factors on the visual information processing of disabled readers*. Unpublished doctoral dissertation, The American University, Washington, D.C.

Carey, J. (1984). *Electronic text and higher education: a summary of research findings and field experiences. Report No. 1 of the electronic text report series*. East Lansing, MI: National Center for Research on Teacher Learning. (ERIC Document Reproduction Service. No. ED257446.).

Cavanaugh, S. (1995). *Digital type design guide: The page designer's guide to working with type*. Indianapolis, IN: Hayden Books.

Chapman, J. C. (1923). *Chapman-Cook speed of reading test*. Ames, IA: Iowa State University Press.

Cisco. (2000). *Cisco systems product trademark terms*. Retrieved May 15, 2000, from http://www.cisco.com/univercd/cc/td/doc/product/lan/c2900xl/gbic/78649405/7864940a.gif

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences: Second Edition*. Hillsdale, NJ: Lawrenece Erlbaum Associates, Publishers.

Colaiuta, W. (2000). *From the Let's-be-like-Sony Dept*. Retrieved August 18, 2000, 2000, from http://wincent.org/showpage.php3?article=opinion/000819aqua.php3

Cross, L. (2000). *Start with Between Subjects First*. Retrieved October 3, 2000

Drewry, H. G. G. (2001). *RE: Professional Printing Standards, e-mail*. Retrieved August 18, 2001

Dyer, B. S. (1997). *The Nelson-Denny Reading Test: An Inside Look*. Retrieved May 19, 2000, from http://www.aug.edu/augusta/students/magazine-march-97/nelson.htm

Electronic Buyer's News. (1999). *New 6.3-in. display from Toshiba boasts print-quality image*. Retrieved November 11, 1999, from http://www.ebnews.com/story/OEG19991111S0007

Eurich, A. C. (1936). *Minnesota speed of reading test for college students*. Minneapolis, MN: University of Minnesota.

Every, D. K. (1999). *Anti-Aliasing: What is it, and how does it work?*, 1999, from http://www.mackido.com/General/AntiAliasing.html

Felici, J. (1996). *A bitmap a day keeps the eye doctor away*. Retrieved November 20, 2000, from http://www.will-harris.com/typofeli.htm

Fields, B. (2000). *Typeface practice onscreen, e-mail*. Retrieved November 22, 2000

Gardner, M. (2000). *Type survey, e-mail response*. Retrieved November 18, 2000

Geske, J. (1996). *A comparison of three type faces for use as body copy in computer mediated communication*.Unpublished manuscript, University of Iowa: Ames, IA.

Geske, J. (2000). *Readability of body text in computer mediated communications: Effects of type family, size and face*. Retrieved October 17, 2000, from http://www.public.iastate.edu/~geske/scholarship.html

Glynn, S. M., & Britton, B. K. (1984). Supporting reader's comprehension through effective text design. *Educational Technology, 24*, 40-43.

Gowen, M. (2000). *Type survey, e-mail response*. Retrieved November 20, 2000

Grant, M. M., & Branch, R. M. (2000). *Performance differences between serif fonts and sans serif fonts in an on-screen reading task*. Paper presented at the International Visual Literacy Association, Ames, Iowa.

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika, 24*(95-112).

Hartley, J., & Burnhill, P. (1977). Understanding instructional texts: Typography, layout, and design. In M. J. A. Howe (Ed.), *Adult learning* (pp. 223-247). London: Wiley.

Hoener, A., Salend, S., & Kay, S. (1997). Creating readable handouts, worksheets, overheads, tests, review materials, study guides, and homework assignments through effective typographic design. *Teaching Exceptional Children, 29*(3), 32-35.

Holmes, J. A. (1986). *Formatting variables and typeface variations of dot-matrix print and their effect on reading comprehension and reading rate*. Unpublished doctoral dissertation, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.

Howell, D. C. (1997). *Statistical Methods for Psychology* (4th ed.). Belmont, CA: Duxbury Press: an Imprint of Wadsworth Publishing Company.

IBM. (2000). *Display technology breakthrough: Researchers develop 200 pixels-per-inch display*. Retrieved May 20, 2000, from http://www.research.ibm.com/news/detail/factsheet200.html

Independent. (2000). *Return of the Mac, new and improved*. Retrieved November 19, 2000, from http://www.independent.co.uk/news/Digital/Features/2000-11/mac131100.shtml

Ipek, I. (1995). *Considerations for CBI screen design with respect to text density levels in content learning from an integrated perspective*. East Lansing, MI: National Center for Research on Teacher Learning. (ERIC Document Reproduction Service. No. ED380080).

Isaacs, G. (1987). Text screen design for computer-assisted learning. *British Journal of Educational Technology, 18*(1), 41-51.

Jones, M. (1994). *Visuals for information access: A new philosophy for screen and interface design*. East Lansing, MI: National Center for Research on Teacher Learning. (ERIC Document Reproduction Service. No. ED380085).

Journa, G. (1989). *Effects of image quality on reading performance and perceived image quality from CRT and hard-copy displays*. Unpublished master's thesis, Virginia Polytechnic Institute and State University, Blacksburg, VA.

Kelsey, A. (2000). *How can I use the Lingo printFrom command? (13054)*. Retrieved November 25, 2000

Keppel, G., & Zedeck, S. (1989). *Data analysis for research designs: Analysis of variance and multiple regression / correlation approaches*. New York: W. H. Freeman.

Kostelnick, C. (1990). The rhetoric of text design in professional communication. *The Technical Writing Teacher, 17*(3), 189-202.

Kramer, R., & Bernhardt, S. A. (1996). Teaching text design. *Technical Communication Quarterly, 5*(1), 35-60.

Landweber, G. D. (2000). *SmoothType 2.2.2*. Retrieved November 18, 2000, from http://www.kaleidoscope.net/greg/smoothtype.html

Luckiesh, M., & Moss, F. K. (1941). The effect of line length on readability. *Journal of Applied Psychology, 25*, 67-75.

Manzel, K., & Tranel, D. (1995). *The development and standardization of a reading test for brain-damaged patients*. Paper presented at the meeting of the Midwestern Psychological Association, Chicago, IL.

Mariano, G. (2000). *Surf's up on the Net—23 percent over last year*. Retrieved November 16, 2000, from http://news.cnet.com/news/0-1005-200-3731909.html

Mason, L. D. (1997). Design issues for producing effective multimedia presentations. *Technical Communication, 44*(1), 65-71.

Maurer, J. (1998). *Technote 1149: Smoothing fonts*. Retrieved December 31, 1999, from http://www.apple.com/support

McLean, R. (1997). *The Thames and Hudson manual of typography*. London: Thames and Hudson Ltd.

Microsoft. (1997). *Font properties extension*. Retrieved May 19, 2000, from http://www.microsoft.com/typography/faq/faq5.htm

Moore, D. M. (1993). *Multimedia: promise, reality & future*. East Lansing, MI: National Center for Research on Teacher Learning. (ERIC Document Reproduction Service No. ED370554).

Mukherjee, P., & Edmonds, G. S. (1993). *Screen design: a review of research*. East Lansing, MI: National Center for Research on Teacher Learning. (ERIC Document Reproduction Service No. ED370561).

Negroponte, N. (1994). *Aliasing: The blind spot of the computer industry*. Retrieved May 20, from http://www.wired.com/wired/archive/2.01/negroponte_pr.html

Ott, L., Larson, R. F., & Mendenhall, W. (1987). *Statistics: A tool for the social sciences* (4th ed.). Boston: PWS-KENT.

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. (1932). Studies of typographical factors influencing speed of reading: X. Style of type face. *Journal of Applied Psychology, 16*, 605-613.

Paterson, D. G., & Tinker, M. A. (1940a). *How to make type readable: A manual for typographers, printers and advertisers*. New York: Harper & Brothers Publishers.

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 size point type. *Journal of Educational Psychology, 33*, 552-555.

Paterson, D. G., & Tinker, M. A. (1946). The relative readability of newsprint and book print. *Journal of Applied Psychology, 31*, 160-163.

Pearson, D. P., Barr, R., Kamil, M. L., & Mosenthal, P. (1984). *Handbook of reading research* (Vol. 1). New York: Longman.

Peck, W. (2000). *Type right for a pro look: Production graphic*. Retrieved November 24, 2000, from http://www.webreference.com/graphics/column26/

Person, R. (2000). *Tests (Test-Tinker, Found at Southern Illinois University), e-mail*. Retrieved June 13, 2000

Phong, A. V. (2000). *Anti-aliasing: Photoshop Tutorial, A Photoshop Anti-Aliasing Tip*. Retrieved November 26, 2000, from http://www.phong.com/tutorials/anti-alias/

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

Riverside-Publishing. (2000). *Nelson-Denny Reading Test: Products and Services*. Retrieved May 19, 2000, from http://www.riverpub.com/products/group/nd_read.htm

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

Rozmiarek, E. J., Burgess, P. S., & Weinfeld, M. N. (1996). *GRE Success*. Princeton, NJ: Peterson's.

Rubin, P. A. (1984). *A primer on electronic text technology for College Administrators. Report No. 4 of the electronic text report series*. East Lansing, MI: National Center for Research on Teacher Learning. (ERIC Document Reproduction Service No. ED259721).

Siracusa, J. (2000). *Mac OS X public beta: Anti-alias in DP 4*. Retrieved October 4, 2000, from http://www.arstechnica.com/reviews/4q00/macosx-pb1/macos-x-beta-9.html

Skordahl, D. M. (1958). Effect of sloping text upon the speed of reading and upon visibility. *University of Minnesota, Unpublished manuscript*.

Smeijers, F. (1996). *Counterpunch: making type in the sixteenth century*. London: Hyden Press.

Starch, D. (1923). Layout and typography. In *Principles of advertising* (pp. 657-669). Chicago: A. W. Shaw.

Streit, L. D., Davis, L. E., & Ladner, C. W. (1986). Structuring text: Creating cost-effective training manuals. *Educational Technology, 26*, 29-31.

Sutherland, S. W. (1989). *Miles Albert Tinker and the Zone of Optimal Typography*. Unpublished Dissertation, University of Washington, Seattle.

Sutherland, S. W. (2000). *Tinker and the legibility of type, e-mail*. Retrieved July 4, 2000, from mailto:IrisPress@aol.com

Taylor, J. L. (1990). *The effect of typeface on reading rates and the typeface preferences of individual readers*. Unpublished doctoral dissertation, Wayne State University, Detroit, MI.

Tinker, M. A. (1939). Illumination standards for effective and comfortable vision. *Journal of Consulting Psychology, 3*, 11-20.

Tinker, M. A. (1956). Effect of sloped text upon the readability of print. *American Journal of Optometry and Archives of the American Academy of Optometry, 33*, 189-195.

Tinker, M. A. (1959). Brightness, contrast, illumination and visual efficiency. *AmericanJournal of Optometry and Archives of the American Academy of Optometry, 36*, 221-235.

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

Tinker, M. A., & Paterson, D. G. (1929). Studies of typographical factors influencing speed of reading: III. Length of line. *Journal of Applied Psychology, 13*, 205-219.

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

Tinker, M. A., & Paterson, D. G. (1944). Wartime changes in newspaper body type. *Journalism Quarterly, 21*, 7-11.

Tullis, T. S., Boynton, J. L., & Hersh, H. (1995). Readability of fonts in the Windows environment. *Association of Computer Machinery, CHI 1995 Proceedings*, 227-235.

Turner, R. J. (1982). *An experimental study of the legibility of minus letterspacing of typography as measured by rate of reading and comprehension*. Unpublished doctoral dissertation, Ohio University.

Vosseller, S. (2000). *Points, pixels, DPI and you*. Retrieved May 2, 2000, from http://www.microsoft.com/mac/IE/Support/article_pixels.asp

Waite, J. (1995). Typographic principles for word processing students. *Business Education Forum, 49*, 28-32.

Weisenmiller, E. M. (1999). *A Study of the Readability of On-screen Text*. Unpublished doctoral dissertation, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.

Weston, H. C. (1935). The relation between illumination and industrial efficiency. In *The effect of size of work* (pp. 14). London.

Weston, H. C. (1945). *The relation between illumination and visual efficiency—the effect of brightness contrast*. London: H. M. Stationery Office.

Williams, R. (1994). *The non-designer's design book*. Berkley, CA: Peachpit Press.

XY_Art. (2000). www.BitchySister.com. Retrieved November 4, 2000, from http://www.bitchysister.com/12.jpg