

Readability of Fonts in the Windows Environment

Thomas S. Tullis, Jennifer L. Boynton, & Harry Hersh

Fidelity Investments

82 Devonshire Street, P10A

Boston, Massachusetts 02109-3614

ABSTRACT

The readability of twelve different fonts and sizes in the Microsoft Windows environment was studied. The specific fonts were Arial, MS Sans Serif, MS Serif, and Small Fonts. Their sizes ranged from 6.0 to 9.75 points. These were presented using black text on either a white or gray background and either bold or non-bold style. There were significant differences between the various font/size combinations in terms of reading speed, accuracy, and subjective preferences. There were no consistent differences as a result of background color or boldness. The most preferred fonts were Arial and MS Sans Serif at 9.75. Most of the fonts from 8.25 to 9.75 performed well in terms of reading speed and accuracy, with the exception of MS Serif at 8.25. Arial at 7.5 and both of the Small Fonts (6.0 and 6.75) should generally be avoided.

KEYWORDS

Font, Text, Readability, Legibility, Windows

INTRODUCTION

The struggle between readability and screen real estate is a frequent problem encountered by Windows application developers. While there have been studies comparing the readability of on-line and printed text, there is no research comparing the readability of fonts for Microsoft Windows applications, which are available in an array of sizes, styles, and colors.

EXPERIMENT

An experiment was conducted to measure readability and preference differences among various Microsoft fonts across selected sizes in which they are available. Twelve specific fonts and sizes, as shown in Table 1, were studied. Each font/size combination was examined as black text on either a white or gray background, and in either bold or non-bold style. This resulted in 48 (12 x 2 x 2) total combinations.

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the ACM copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of ACM. To copy otherwise, or to republish, requires a fee and/or specific permission.
CHI' Companion 95, Denver, Colorado, USA
© 1995 ACM 0-89791-755-3/95/0005...\$3.50

Font / Size	
Small Font	6.0
Small Font	6.75
Arial	7.5
Arial	8.25
Arial	9.0
Arial	9.75
MS Sans Serif	8.25
MS Sans Serif	9.75
MS Serif	6.0
MS Serif	6.75
MS Serif	8.25
MS Serif	9.75

Table 1. Twelve font and size combinations studied.

Subjects

Fifteen volunteer subjects participated in the experiment, ranging between 27 and 45 years of age.

Equipment & Environment

The entire experiment was controlled by two programs written in Microsoft Visual Basic. The programs were run on a 486/33MHz PC (8 Mb of RAM) and a NEC 5FG 15" monitor running in 1024 x 768 (Small Fonts) resolution.

Procedure

Subjects were given instructions and two practice trials to familiarize them with their task. Each of the 48 font combinations was represented as one trial, resulting in 48 trials generated in random order. Each trial began with a dialog box instructing subjects to count the number of typographical errors in the paragraph to follow. Subjects were instructed to press the enter key, read through the paragraph once, and count the typographical errors as they read. When the subjects were done reading, they were instructed to press the enter key again, which brought up a dialog box allowing them to input the number of errors. After entering a number, the next trial began. Search time was measured from the time the paragraph appeared on the screen until the enter key was pressed.

The same paragraph appeared throughout all trials in the same position on the screen. The typographical errors were randomly generated; however, there were always between 1 and 5 errors in a paragraph. An error consisted of a randomly selected letter being replaced by a different random letter.

After completing the 48 trials, subjects were instructed to perform a preference task, which consisted of rating each of the 48 font combinations. The word 'Example' appeared on the screen 48 times, in each of the 48 combinations. Subjects were asked to drag and drop each of the examples into one of four boxes corresponding to a four-point scale of legibility: Poor, Fair, Good, and Excellent.

RESULTS

Three types of data were submitted to analyses: reading time, accuracy, and preference. A within-subjects analysis of variance was conducted for each measure. Font/size was treated as one variable which was orthogonal to the other two variables (background color and boldness).

Time Data

Analysis of the reading times, as shown in Figure 1, revealed a significant main effect of font/size, $F(11,154) = 4.15, p < .0001$. None of the other main effects (color and boldness) were significant, nor were any of the interactions. Post-hoc comparisons of means revealed the following:

- "Arial 7.5" and "Small Fonts 6.0" were significantly worse than all others.
- "MS Serif 8.25" was significantly worse than "Arial 9.75", "MS Sans Serif 8.25", and "MS Serif 9.75".
- "MS Serif 6.0" and "MS Serif 6.75" were significantly worse than "MS Serif 9.75".

Accuracy Data

Accuracy was determined by whether or not the subject reported the correct number of typographical errors for the trial. Analysis of the accuracy data, as shown in Figure 2, revealed a significant main effect of font/size, $F(11,154) = 3.71, p < .0001$. None of the other main effects were significant; however, there was a significant interaction between font/size and boldness. Post-hoc comparisons of means revealed the following:

- "Arial 9.75" was significantly better than "Small Fonts 6.0", "Arial 7.5", "Small Fonts 6.75", and "MS Serif 8.25".
- "Arial 9.0" and "MS Sans Serif 9.75" were significantly better than "Small Fonts 6.0" and "Arial 7.5".
- "Small Fonts 6.0" was significantly worse than all others.

Preference Data

Preferences were coded on a four-point scale, where 1=Poor and 4=Excellent. Analysis of these ratings, as shown in Figure 3, revealed a significant main effect of font/size, $F(11,154) = 3.71, p < .0001$. None of the other main effects were significant; however, the interaction between font/size and boldness was significant. Post-hoc comparisons of means revealed that most of the means shown in Figure 3 were significantly different, with the following exceptions:

- The two best fonts, "MS Sans Serif 9.75" and "Arial 9.75" were not significantly different from each other.
- "MS Serif 9.75" and "MS Sans Serif 8.25" were not significantly different from each other.

- "MS Serif 8.25", "Arial 7.5", "MS Serif 6.0", and "MS Serif 6.75" were not significantly different.

CONCLUSIONS

Obviously, there were distinct differences in the speed and accuracy with which users could read these various fonts, and even stronger differences in their subjective preferences for the fonts. However, there were no consistent differences for white vs. gray backgrounds or bold vs. non-bold styles.

Looking at the data as a whole, a few conclusions appear warranted:

- "Arial 7.5" and both "Small Fonts" should generally be avoided.
- To optimize subjective preference, use "Arial 9.75" or "MS Sans Serif 9.75".
- To optimize reading speed and accuracy, the best choices appear to be any of the fonts at 8.25, 9.0, or 9.75, except for "MS Serif 8.25".

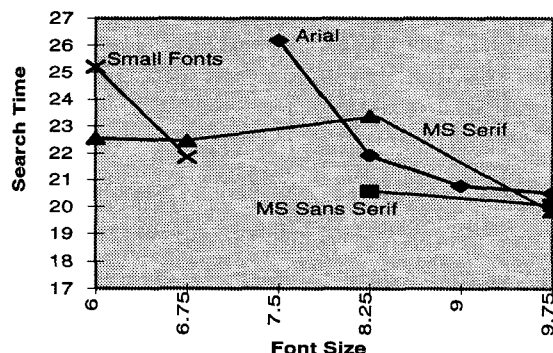


Figure 1. Reading time (sec) for fonts as a function of size.

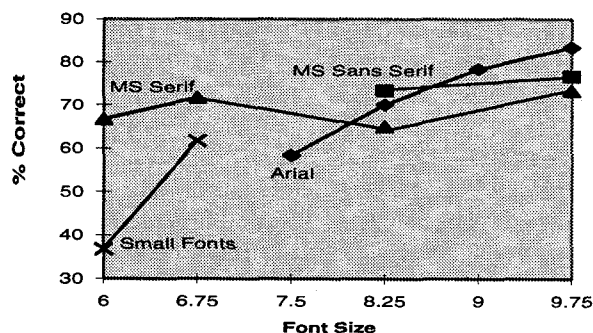


Figure 2. Percent correct for fonts as a function of size.

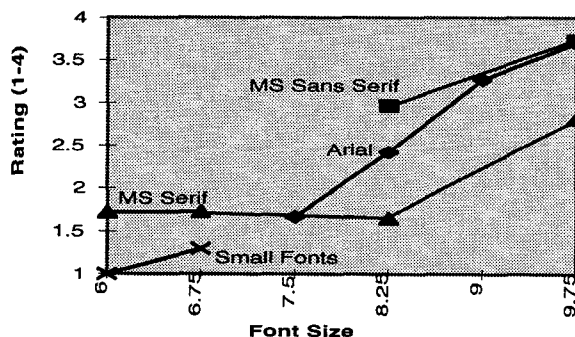


Figure 3. Subjective ratings of fonts as a function of size.