AMERICAN JOURNAL OF PSYCHOLOGICAL RESEARCH

Volume 1, Number 1
Submitted: January 10, 2005
First Revision: May 1, 2005
Accepted: May 13, 2005
Publication Date: May 20, 2005

Effects of Typeface and Font Size on Legibility for Children

Rebecca J. Woods, Kristi Davis, and Lauren F. V. Scharff
Stephen F. Austin State University, lscharff@sfasu.edu

ABSTRACT

Eighty children (kindergarten through 4th grade) performed discrimination and identification tasks for tachistoscopically-presented letter pairs. Letter pairs manipulated font (Times New Roman, Arial), point size (12- and 18-point) and letter-pair discrimination difficulty (easy, medium, and hard). Higher percent correct scores occurred for the Arial font and 18-pt size, and performance was better for discrimination than identification. Effects and within-grade variability were larger for the lower grades, with performance differences decreasing by the second grade. Letter frequency better predicted letter-pair discriminability for the lower grades, but by the fourth grade, the trends began to match adult discriminability categories. These differences may be due to both physiological (e.g., eye movements) and cognitive (e.g., memory and attention) development.

INTRODUCTION

Much research has been dedicated to the effects of typeface and font size on legibility and readability of text in adult populations. Researchers have applied little of this information, however, to empirical studies of these topics in children. It is important to examine children as a specific population due to the effects that developmental changes and familiarity with characters (or the lack of) may have on the readability of text. Because a large majority of children's reading material (i.e., textbooks, standardized tests, literature) is found in printed form, it is particularly important to examine readability of printed text for children. Publishing companies have guidelines, but these are often based on font types and sizes most frequently used by other publishing companies rather than on empirical data investigating legibility and readability (L. Gerbrandt, personal communication, May 16, 2000).

When one examines readability, the ease (speed and comfort) with which an individual reads and comprehends text, it is important to first determine legibility, the ability of an individual to distinguish characters. The legibility of text influences its readability (Erdmann & Neal, 1968), which in turn influences ease of comprehension of written material. For this reason, the current study examined developmental changes in children's ability to discern letters of individual fonts, rather than investigating readability of text. We hoped that this study will serve as a first step in an effort to resolve questions regarding what are the most appropriate font types for children's reading material.

There are several reasons to suspect that differences will exist between legibility and readability of letters for children compared to adults. Particularly relevant to studies of legibility and readability in children are the effects of physiological development. At this time the precise nature of this influence is uncertain. Research indicates that the human eye is fully developed after the first year (Slater, 1998) and, children, like adults, are able to discriminate small visual details (Gaines, 1969). This suggests that even subtle differences in text characteristics, such as typeface or size, should also be distinguishable by children. However, limitations may be imposed at higher processing levels because cortical areas involved in reading continue to develop through adulthood (Pinel, 2000).

Similarly, inefficient occulomotor patterns may also affect children's ability to distinguish and read letters effectively. Infants' occulomotor patterns are considerably less efficient than adult patterns (Aslin, 1985, 1987). Despite rapid development in the efficiency of eye movements, by the time children enter school eye-movement patterns remain significantly less efficient than those of adults (Kowler & Martins, 1982). According to Tinker (1963), it is not until the fourth grade that legibility for children parallels that of adults because it is at that age that occulomotor patterns stabilize. However, Krueger (1973) found that although children in fourth grade appear as responsive as adults on high-level factors such as orientation and frequency of letter occurrence, children process letter information more slowly than adults. This may also be due to immature cortical areas. The extent to which these changes may affect legibility remains unclear. These studies highlight the importance of investigating younger populations, particularly those prior to the fourth grade for whom it is clear that immaturities in occulomotor patterns and cortical development exist.

Despite this need, few studies have focused on legibility of letters for children; consequently we draw largely on information from adult studies. These studies suggest that many factors influence legibility. These include, 1) font or typeface characteristics, such as stroke width or letter size (Erdmann & Neal, 1968; Jha & Daftuar, 1981; Krulee & Novy, 1986; Mansfield, Legge, & Bane, 1996; Roethlein, 1912; Sanocki, 1988; Smith, 1979; Tinker, 1944, 1959; Yager, Aquilante, & Plass, 1998), and 2) presentation factors, such as luminance contrast or resolution (Charness & Dijkstra, 1999; Erdmann & Neal, 1968). We discuss these factors further in the following section. Legibility differences also exist between types of stimuli, for example, compared to printed material, computerized presentations limit on-screen reading performance (Gould, et al., 1987; Kruk & Muter, 1984; Mills & Weldon, 1987). Vision problems, such as myopia or astigmatism, also influence legibility but are beyond the scope of this study.

Font Characteristics

Many studies have examined the effects of variations in typeface characteristics on legibility and readability. Differences in font increase or decrease the ability of an individual to distinguish or read letters (Tinker, 1944). According to Chauncey (1986), different fonts may influence reading speed as much as 30 percent. When well above recognition threshold, font differences (e.g., stroke width, x-height) are less influential on legibility (Tinker, 1944). However, under certain conditions, variations in font characteristics may particularly influence the legibility of letters. For example, legibility may be compromised due to presentation factors, such as when one uses faster rates of presentation, smaller letters, greater distances, poorer graphic resolutions or contrasts, or inadequate illumination (e.g., Legge, Rubin, & Luebker, 1987; Mansfield, et al., 1996). It is under these conditions that variations in type should make the greatest differences in legibility. Similarly, when any factor stresses the visual system, differences in font greatly affect legibility (e.g., Yager, Aquilante, & Plass, 1998).

Different fonts result from combinations of a number of font characteristics, including serifs, stroke width, and letter height and width (Lannon, 2000). One may study legibility or readability by examining individual characteristics of a font or by studying differences among whole fonts. Both approaches have limitations and may lead to different conclusions. Systematically studying specific characteristics of a font, such as stroke width or the use of serifs, requires manipulation of individual characteristics while holding others constant. While this approach allows direct assessment of the influences specific font characteristics have on legibility, it often proves impractical because it may require construction of artificial fonts. Conversely, studying whole fonts, such as Times or Helvetica, though easily used in practical application, limits generalizeability with respect to specific characteristics; one can generalize results only to fonts closely resembling the chosen font. Table 1 provides an example of the two fonts we used in the current study and how they vary on some of these specific characteristics.

Table 1
Font differences between Times New Roman and Arial.

Font Characteristics	Arial	Times New Roman
Serifs	No	Yes
Stroke Width	Uniform	Variable
Letter Height	Larger	Smaller

Serifs are the fine lines that extend horizontally from the main strokes of a letter, as seen in Times New Roman (Lannon, 2000). Fonts are referred to as sans serif fonts when they do not include extra horizontal strokes, as seen in the font Arial. The purpose of serifs, according to Lannon, is to guide the reader's eye horizontally, thereby making the type more readable. However, the effectiveness of serifs remains controversial. For example, several studies investigating the readability of fonts suggest that Lannon's assumption may be incorrect. These studies reported that sans serif fonts are more readable than serif fonts; nevertheless results are disputable. One reason is that in these studies researchers did not hold the other font characteristics constant. For example, Yager, Aquilante, and Plass (1998) reported that participants preferred a sans serif font to a serif font for readability, but the researchers had failed to control stroke width. Similarly, Krulee and Novy (1986) found that participants preferred a sans serif font over other measured fonts; however, the measured fonts varied in several other ways including stroke width, size and ornamentation. Sanocki (1988) also found a sans serif font preferable to a serif font. Researchers may need to further modify these conclusions to take into account a study by Jha and Daftuar (1981) that suggests that serifs are preferable at larger letter sizes, while sans serifs are better at smaller letter sizes. In effect, these studies fail to provide conclusive evidence of the influence of serifs on legibility or readability. Other components of font may prove more influential.

For example, stroke width also appears to influence legibility (Krulee & Novy, 1986; Roethlein, 1912; Yager, Aquilante, & Plass, 1998). Stroke width refers to the width of each component or stroke of a letter. A font may have a uniform stroke width in which all of each stoke is the same width, as seen in the font Arial, or a font may have a varying stroke width in which some portions of the stroke are thinner than others, as seen in the font Times New Roman. A font may also have a generally wide stroke width, referred to as "bold", or a thin stroke width, referred to as "light." Letters with uniform stroke width appear to be more legible (Yager, Aquilante, & Plass, 1998) as do letters that are bold (Krulee & Novy, 1986). Serif fonts typically have variations in stroke width whereas san serif fonts typically have uniform stroke width (Anderson, 1987).

Letter height, or x-height, is the specific height of the lower case letters of a font. Letter height is different from point size, as letter height can vary between fonts of the same point size. For example, the x-height for Times New Roman at 12-points is smaller than Arial at 12-points. Researchers have found that both letter height and point size influence legibility. In general, larger letters provide better legibility (Erdman & Neal, 1968; Roethlein, 1912; Smith, 1979), however, only to an upper threshold beyond which legibility decreases as size increases (Legge, et al., 1987). This effect also occurs for measures of readability, even for children (e.g., Hughes & Wilkins, 2002).

Presentation Characteristics

The manner by which one presents letters can also affect legibility and readability (Legge, Pelli, Rubin, & Schleske, 1985). For example, reducing reading distance or increasing letter height increases the size of the retinal image created by a letter, thus improving legibility. Other presentation factors such as letter resolution and case may also have profound effects on legibility.

Because the quality of letter resolution in a display can increase or decrease legibility, a disparity exists between the legibility of printed text compared to computer-displayed text (e.g., Gould et al., 1987; Mansfield, et al., 1996). Higher resolution of letters results in better legibility; therefore due to limitations of letter resolution on computer screens (e.g., 72 DPI), printed text (i.e., 300 DPI) is more legible. Research has yet to determine the extent that such resolution differences affect legibility in children. For this reason, we chose to assess legibility using a method (tachistoscopically-presented slides of printed text) that will allow resolution equitable to printed text rather than assessing legibility of computer-displayed text. For a related study using computerized stimuli in children older than those in the current study see Bernard and colleagues (2002).

Differences also exist when presentation consists of upper or lower-case letters for both legibility and readability. According to Tinker (1963), upper-case letters are more legible than lower-case letters when presented individually. This is probably because upper-case letters contain more horizontal and vertical strokes, which are easily processed, as opposed to lower-case letters that have more curved strokes. Upper-case letters also have more open spaces between strokes than lower-case letters allowing for better discrimination between letters. However, upper-case letters are less readable when presented as words or strings of letters. Tinker's research has shown that the word forms created by using upper and lower case letters combined is beneficial in the reading process. The use of only upper case letters removes the word forms from text thereby hindering readability.

The current study investigated developmental aspects of legibility in elementary-school age children (kindergarten to fourth grade). We compared the legibility of lower-case letter pairs in two fonts, Times New Roman and Arial, for two type sizes, 18-point and 12-point. These correspond to font types and sizes that typically appear in printed publication for children across these grades (L. Gerbrandt, personal communication, May 16, 2000). We chose letter pairs to specifically produce easy, medium, and hard levels of distinguishability. We based these categories on the adult data (Sanford, 1888) and we will refer to these as letter-difficulty levels.

Legibility research frequently examines distinguishability of individual letters, rather than examining letters grouped as words or non-words. Letter-difficulty levels come from several studies that have examined the relative legibility of letters of the English alphabet (Bouma 1971; Roethlein 1912; Sanford, 1888). The letters from these studies were grouped into three categories by their ability to be distinguished: easy (high legibility), medium (medium legibility), and hard (low legibility). Similar results are obtained in research on confusion matrices (Gervaise, Harvey, & Roberts, 1984) which determine which letters are most easily confused with other specific letters.

Our study used two tasks to assess legibility. One task (discrimination) required participants to make same-different judgments of letter pairs thus indicating that the letters were distinguishable. A second task required that participants identify each letter in a pair. We used these two tasks so that potential performance differences due to higher-level “alphabet learning” could be differentiated from performance differences due to letter legibility.

We expected Arial to have higher measures of legibility due to its uniform stroke width and larger letter height. Of the two type sizes used, we expected that the larger type size would be more legible. Legibility measures were expected to increase monotonically with grade due to increasingly efficient occulomotor patterns, improved cognitive processing, and reading experience. We expected results to follow the letter-difficulty levels of easy, medium, and hard. Finally, we expected same-different judgments (discrimination task), to have better overall performance and to be less influenced by other variables than a letter identification task because same-different judgments require fewer processing demands.

METHOD

Participants

We tested a total of 91 participants. Five of the participants came from through the University Learning Center, and 86 participants were recruited through the Charter School on the university campus. Participants were enrolled in classes ranging from kindergarten (K) through fourth grade. The number of participants in each grade was as follows: K = 23, 1^{st} = 20, 2^{nd} = 16, 3^{rd} = 17, and 4^{th} = 15. All participants had a signed parental consent form on which their parents indicated normal or corrected-to-normal vision of the participants. Compensation for participation included a cup and sticker with the university’s emblem.

Design and Stimuli

The experiment was a 2(task) x 2(font) x 2(size) x 3(letter-difficulty level) x 5(grade) mixed design. We used two different tasks, discrimination and identification, to test legibility of two different fonts, Times New Roman (TNR) and Arial. An 18-point and a 12-point font size were tested for both fonts. All participants, ranging in grade from kindergarten to fourth, were tested with the two fonts, two type sizes, and three letter-difficulty levels for both tasks.

We used a discrimination task and an identification task to determine the legibility of the letter pairs. The discrimination task required determining if letter pairs consisted of the same two letters (e.g., ss) or two different letters (e.g., se). Skills necessary to perform the discrimination task were limited to visually recognizing the difference between two letters and indicating so on the answer sheet. The identification task required writing the letter pairs; some knowledge of the letters and experience with writing letters were necessary for this task.

The two fonts tested differed in their specific characteristics as shown in Table 1. Arial is a san serif font with uniform stroke width, and TNR is a serif font with variable stroke width. Arial’s letter height is slightly larger than TNR in the same point size. These fonts and sizes were chosen in response to information from publishing company guidelines. Variations in serifs were based on commonly-used styles in printed texts for children in kindergarten through fourth grades. These guidelines also show that 18-point fonts are commonly used for kindergarten texts, and point size decreases to 12-point by about third grade (L. Gerbrandt, personal communication, May 16, 2000).

Letter-pair choices for this study were based on letter recognition information from Sanford's (1888) order of recognition list based on adult data. All letter pairs were in lower case. The letter pairs were m and w, b and h, and e and s. Each of the letters in a pair had equal difficulty ranking in Sanford's letter recognition list. The easily recognizable letters were m and w, and the medium letter-difficulty letters were b and h. The letters e and s fall into the difficult-to-recognize category and are frequently confused with each other.

Apparatus and Materials

Letter pair displays were presented using a Kodak Ektographic III slide projector with 41010-A Projection Tachistoscope. Slides were produced by photographing printed text onto slide film using a Canon Rebel 35mm camera. A total of 99 slides were produced. Three practice trial slides were made using non-test letter pairs. Test slides consisted of each letter pair in every possible order (e.g., mm, mw, wm, ww) and these pairs were repeated so there were four trials per condition.

A quick vision test was created with printed text. The vision test consisted of six lines of letters, one line each in 10, 12, 14, 16, 18, and 20-point type sizes with the largest letters at the top, similar to a Snellen Chart. This range of type sizes was used to encompass the font sizes of the letters used in this study. Each line contained three letters, one each from the easy, medium, and hard letter-difficulty levels (from top to bottom: B D E, R Q C, L T V, M P K, F H O, X U A).

Responses were recorded by participants on answer sheets. Pencils and clipboards were supplied to the participants for answering. The discrimination task for each trial required the participant to circle an S if the letters were the same and a D if the letters were different. The participants fulfilled the identification task by writing the letter pairs on a blank line directly to the right of the discrimination response area.

Additional Controls

All participants from the Charter School were tested in the same room, which had minimal lighting from interior skylights. The Learning Center participants were tested in a different room with no windows. During the testing, all room lights were off so that the only lighting was from the projector. The reflected luminance at the participants' viewing distance was 25 cd/m^2.

We chose an exposure time so that the participants could view the stimuli just above the lower thresholds of time needed for accurate visual and mental processing. The exposure interval was set at .2 seconds. Pre-tests were performed in which exposure times both faster and slower than .2 seconds were tested on six children to determine the fastest interval which was still within the limits of the children's capabilities.

Due to the quick stimulus presentation, we needed to specify the target area to the participants to avoid missed trials. We placed a target square made of light yellow paper, which measured 4.1 inches by 2.5 inches, in the viewing area. This enabled the participants to focus on the proper place at the time of display.

Rooms were arranged identically for each group, with particular attention given to the visual angle of the stimuli. Each room was arranged to create the visual angle necessary to make the display roughly equivalent to 12-point and 18-point printed text at an average reading distance of 18 inches. The 18-inch reading distance was based on a reading position where a book was placed on a table surface versus being held. For a TNR 12-point lower case "m" the visual angle was .004 at the 18-inch reading distance. In order to match this with the slides, the target box display area was placed 3 ft. 8 in. from the floor. The projector was positioned 8 ft. 1 in. from the screen, and the viewing distance was 6 ft. from the screen (chairs were positioned so the average child's head and eyes would be 6 ft from the target box). Participants' actual viewing position within the chair was not restrained, so there was some variance in actual viewing distance; however, this variance should be analogous to the natural variance that occurs when seated and reading printed text from a book.

Procedure

All participants were recruited by their teachers using the experiment proposal information and a parental consent form. Test groups ranged in size from one participant to four participants depending upon participant availability at the testing time.

Each session began with standardized instructions to explain the general nature of the test. The quick vision test was then administered during which the participants wrote each of the letters of the vision test into a provided space on the answer sheet. The discrimination and identification tasks were then explained in more detail. Examples were shown, and the concepts of "same" and "different" letters were clarified. The participants were then told how to mark their answer sheet. For the discrimination task, the participants were instructed to circle the S if the letters shown were the same and to circle the D if the letter in a pair were different. For the identification task, the participants were instructed to write the letters they had just seen in the correct order (e.g., if the trial letter pair was es, then se would be considered incorrect). Order of task was counterbalanced between groups to control for order effects by instructing half of the participants to write the letters first and the other half to circle S or D first. Prompting from the researcher was continued throughout all trials to ensure that the participants were looking at the target screen and to control order of task completion. However, many students were observed to consistently complete the identification task first regardless of prompting.

Three practice trials were shown first and answers were checked to make sure the participants knew how to complete the task. We instructed participants to skip any trials in which they could not see the display for any reason. Participants took at least one break during each session (younger participants often needed more breaks than the older participants). Ninety-six real trials were shown to each group of participants. These real trials appeared in a pseudorandom order. We initially randomized the slides prior to placement in the projector carousel and we counterbalanced the order of slide presentation between groups. Half of the sessions viewed the trials in the original order within the carousel, and half of the sessions viewed the trials in the reverse order.

The average time to complete the sessions varied across the grades. Third and fourth graders completed their sessions the fastest, with average times of 26 minutes and 30 minutes respectively. First graders averaged 36 minutes to complete the session, and second graders averaged 39 minutes. Kindergarteners averaged 48 minutes to complete a session (All session lengths include instruction as well as trials). Two groups of kindergarteners did not complete the identification task due to their extreme difficulty with the task and time constraints. The extended time needed for kindergarteners to complete sessions appeared to be due to an earlier onset of fatigue than was seen in the higher grades. Kindergarteners also needed breaks more often during each session in order to stay focused.

In addition, cheating occurred in all grades. We noted that some participants, when failing to properly see a slide, were taking answers from other participants' answer sheets. We removed these participants' answers from analysis.

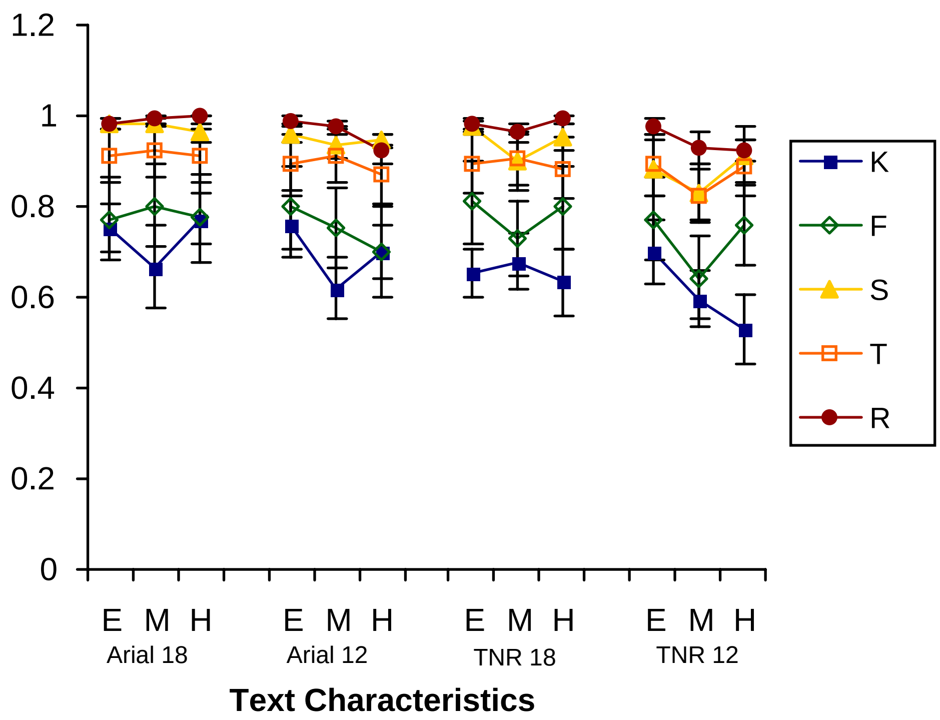
RESULTS

Due to cheating by participants and pre-test runs, we excluded data from 11 participants from analysis. The number of participants included in the analysis was: 17 kindergarteners, 15 first graders, 16 second graders, 17 third graders, and 15 fourth graders. The percent correct scores for each task were tabulated for each participant and analyzed using a 2(Task) x 2(Font) x 2(Size) x 3(Letter-difficulty level) x 5(Grade) mixed analysis of variance (ANOVA). This analysis showed that performance detriments occurred most frequently for the more difficult conditions for the lower grades.

We found significant main effects for all variables. The san serif font, Arial, showed higher scores than the serif font, TNR, $F(1, 75) = 14.92$, $p < .001$, with average scores of 83% and 79% correct, respectively. Higher scores were consistently found for the 18-point font than the 12-point font, $F(1, 75) = 24.99$, $p < .001$, with average scores of 83% and 79% correct, respectively. These results are consistent with research on adult participants.

Figures 1 and 2 (discrimination and identification tasks, respectively) show mean performance scores for each grade in each condition. As seen by comparing Figures 1 and 2, overall percent correct scores for the discrimination task were higher than for the identification task, $F(1, 75) = 7.73$, $p < .01$; the average discrimination task score was 85%, and the average identification task score was 77%. These graphs also illustrate that kindergarteners and first graders scored significantly lower on both tasks than all other grades, but they were not significantly different from each other, $F(4, 75) = 31.94$, $p < .001$. We found no significant differences on task performance between second, third, and fourth graders.

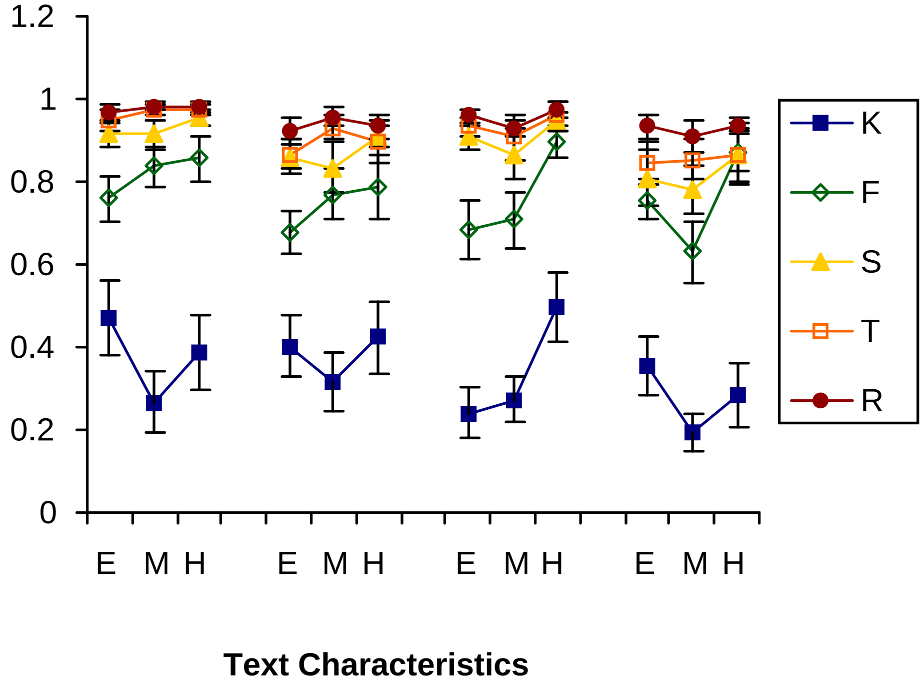
Figure 1: Percent Correct for Discrimination Data as a Function of Text Characteristics (Font, Size) and Discrimination Difficulty (Easy, Medium, or Hard, Based on Adult Data).



Finally, letter-difficulty levels significantly influenced performance, $F(2, 150) = 6.40$, $p < .01$), with the medium letter-difficulty level (bh) having lower scores than the easy (mw) and hard (es) letter-difficulty levels. The easy and hard letter-difficulty levels were not significantly different from each other.

There were also several significant interactions which modify the simple main effects. There was an interaction between grade and task, $F(4, 75) = 6.12$, $p < .001$. The more difficult task, identification, had a more detrimental effect on performance for the lower grades. Kindergarteners scored worse on the identification task than all other grades for either task. Both kindergarteners and first graders scored lower on the discrimination task than did the fourth graders. In addition, as shown by the standard error bars on Figures 1 and 2, the variance within kindergarten and first grades was larger than in any other grade. A comparison of the discrimination data (Figure 1) and the identification data (Figure 2) shows that the kindergarteners' and first graders' significantly poorer performance on the identification task contributes greatly to the overall difference between tasks (notice that no apparent difference exists between tasks for the higher grades).

Figure 2: Percent Correct for Identification Data as a Function of Text Characteristics (Font, Size) and Discrimination Difficulty (Easy, Medium, or Hard, Based on Adult Data).



There was also an interaction for font and letter-difficulty level, $F(2, 150) = 4.30$, $p < .05$. Medium letter-difficulty level for TNR showed lower scores than all other levels. No significant difference occurred for the letter-difficulty level in Arial. Also, letter-difficulty level interacted with task to affect performance, $F(2, 150) = 13.63$, $p < .001$. For the discrimination task, the easy letter-difficulty level letters had higher scores than the hard, with the medium difficulty letters leading to the worst performance. However, for the identification task, the easy letter-difficulty letters had lower scores than the hard, but again, the medium difficulty letters led to the worst performance.

There were two three-way interactions in this analysis. An interaction between grade, task, and letter-difficulty level occurred, $F(8,150) = 1.94$, $p < .06$. A trend indicated lower scores for the kindergarteners and first graders in the identification task for the medium letter-difficulty level. Also, a significant interaction between grade, font, and letter-difficulty level was obtained, $F(8,150) = 2.399$, $p < .05$. This interaction shows that the medium letter-difficulty level (bh) improves after first grade for the font Arial; however, the medium letter-difficulty level (bh) maintains consistently lower scores across grades than the other letter-difficulty levels in the font TNR.

Finally, a significant four way interaction between font, size, task, and letter-difficulty level, $F(2, 150) = 4.87$, $p < .01$) follows the same trends as seen for the main effects; for both tasks at the medium letter-difficulty level, TNR, 12pt., scores are especially low compared to other conditions.

DISCUSSION

All hypotheses were supported and some additional interactions were obtained. As was expected, Arial was found to be more legible than TNR across all grades. Previous research on adults provides several explanations for these results. TNR, like most serif fonts, has variable stroke width (Anderson, 1987), which has been shown to have lower measures of legibility (Yager, Aquilante, & Plass, 1998). In addition, we suggest that serifs have the potential to act as visual noise, cluttering incoming visual information. This gives Arial an advantage over TNR because Arial has uniform stroke width and no serifs. Also, Arial's x-height is larger than TNR in the same point size giving Arial a size advantage. Presently san serif fonts are commonly used when presenting text and equations for mathematical concepts in printed educational materials for children. In contrast, serif fonts are more commonly used in printed presentations of children's reading materials (L. Gerbrandt, personal communication, May 16, 2000). Sanocki (1988) and Tinker (1944) state that familiar fonts are more likely to have better legibility. Although children are exposed to both fonts, the serif font TNR (used for reading material) is seen more often, thereby making it more familiar than the sans serif font, Arial. Nevertheless, Arial appears to be more legible. This suggests that, at least for simple letter-pair comparisons and identification, Arial's font characteristics give it an advantage over TNR that outweigh the advantages TNR gains through familiarity. Consequently, we suggest that use of a san serif font similar to Arial may be particularly helpful to early readers (Kindergarten and first-grade reading levels) by easing letter discrimination during letter learning (Liberman, Shankweiler, & Liberman, 1989; Bachmann, 1991). However, considering that fonts similar to TNR are so often used for adult reading material there is also some need for exposure to TNR as reading skills broaden.

Results from this study also correspond to adult studies examining the effect of point size on legibility. Here the larger point size (18-point) was found to be more legible for kindergarteners and first graders than the smaller point size (12-point). However, after the first grade, the point sizes used in this study showed no significant influence on performance. This indicates that the two sizes used in this study were equally legible for grades two through four, which supports the use of fonts as small as 12-points beginning at approximately the second grade.

Although there was little difference in task performance for second through fourth grades, task performance of lower grades was significantly poorer than higher grade levels. Kindergarten and first grade students had lower scores on both tasks than other grade levels, although more so for the identification task than the discrimination. It is likely that this pattern is due, to some extent, to reduced ability to discriminate and recognize letters at these ages. In addition, our study found that participants scored better on the discrimination task than the identification task for all conditions. This supports Tinker's (1944; 1963) assertions that readability requires more than simply a high level of legibility.

Poorer performance in the two lowest grades may have been affected by immature cognitive processing, including memory and attention. Even though Kindergarten and 1st grade age children may be able to see and identify correctly, they can be expected to have a more difficult time holding letters in their memory long enough to report an answer. This may also explain why kindergarteners scored significantly lower than all other grades on the identification task. For the identification task, the children had to recognize the letters, mentally label them, hold them in memory, and, using writing skills, reproduce them. The information needed to answer correctly on the discrimination task was less complicated and would therefore be easier to hold in memory until a response was given. Because the two tasks were paired, the increased concentration necessary to perform the identification task may have lowered scores on the discrimination task and produced overall lower scores for both tasks for the two lower grades.

Attention span may also have influenced performance on the two tasks, particularly for children in kindergarten and first grade. While a decrease in attention and an increase in fatigue were observed across grades, they were exaggerated and occurred earlier in the session for the two lowest grades. Our pseudorandom trial order will have helped minimize the impact of these effects.

However, it is also likely that other variables affected overall performance and, therefore, demand attention. For example, kindergarteners and first graders are less familiar with letters than children in higher grades. Children in the two lower grades have less experience recognizing and writing letters and are also less familiar with test-taking. Although first grade marks the first mandatory entry into school, most children also attend kindergarten. As a consequence much more variability exists in students' letter writing ability during early schooling; some kindergarteners have been in pre-school or have parents who provided early reading experience, while others have had no experience reading or writing letters. Furthermore, because this experiment took place relatively early in the school year (primarily October), children who had minimal experience with the alphabet had little time to become familiar with recognizing or writing letters, to practice writing, or to gain the experience needed to refine the fine motor skills used for holding a pencil and effectively reproducing a letter. First graders, on the other hand, are more familiar with recognizing and writing letters than are Kindergarteners. Even so, many are still not reading easily. This may explain why the identification task was more difficult for first graders than it was for the higher grades. We suggest that for children in Kindergarten and first grades, future studies will benefit by the implementation of additional practice sessions for young children.

Our data show that the medium letter-difficulty pairs (bh) had significantly lower scores than either the easy letter-difficulty pairs (mw) or the hard letter-difficulty pairs (es). These results are inconsistent with Sanford's (1888) findings for letter recognition in adults. This inconsistency may have been due to letter familiarity. Letters that are used more frequently are more familiar and are, therefore, more easily recognized (Bouma, 1971). Letter frequency information shows that b and h are the least frequently used letters of our letter pairs, followed by m and w, while e and s are the most frequently used letters (Hanna, Hodges, & Hanna, 1971). Our results better correlate with frequency of use than with Sanford's recognition difficulty information.

However, in support of Sanford's (1888) list using adult participants, the fourth graders' scores on the discrimination task for the 12-point fonts showed the monotonic decrease in scores with easy, medium, and hard letter-difficulty level as expected. This may be due to increased familiarity with all letters by the time a student reaches the fourth grade, so the infrequency of b and h did not ensure unfamiliarity. In addition, when comparing the easy and hard letter-difficulty levels for the discrimination task, the easy letter-difficulty level consistently showed higher scores than the hard letter-difficulty level, which supports Sanford's list for adult data. These results may be due to the lower level of processing required for the discrimination task; letter frequency may not have influenced the discrimination task as strongly as the identification task.

The cognitive processes involved in performing the discrimination and identification tasks are not clearly separated. The processes of discriminating, recognizing, and then identifying a letter pair occur along a continuum (Kaye, Brown, Post, and Plude, 1981) with higher levels requiring memory and effort. In an effort to separate the cognitive processes involved (discriminating then recognizing and identifying) prompting was used. An experimenter said, "Circle" for the discrimination task and, "Write" for the identification task. Regardless of prompting order, the participants more often performed the identification task first. We suspect that this was a strategy adopted to maximize "correct" answers. We presume participants wanted to "dump" letter information before it was lost from short-term memory. If participants chose to complete the discrimination task first, they risked losing the information for the identification task while assessing the discrimination task. However, by completing the identification task first, the information from the completed identification task could then be used to complete the discrimination task, thereby simplifying the process. This strategy likely enabled the students to obtain correct answers on both the identification and the discrimination task. Because the ability to identify is contingent on the ability discriminate, we feel confident that despite the use of this particular strategy, our results accurately reflect the ability to distinguish and identify the letters.

In summary, our study demonstrates that, similar to results from adult studies, large point sizes are more legible for young children. While this pattern was not obtained for children in grades 2 through 4 using the point sizes studied here, younger children did perform significantly better when viewing the larger of the two point sizes. Also of significant interest is that Arial, a sans-serif font, was shown to be more legible than Times New Roman, a serif font, for all grades. The implications of this finding are of particular interest, given the ongoing debate concerning the "better" font style: a serif or sans serif. Our study supports the use of a sans serif font for young children.

It is our hope that the results of this study on legibility provide a foundation for further studies on both legibility and readability for children. Because readability encompasses other factors, such as leading and page layout (Tinker, 1963) we suggest that additional studies must be conducted before final recommendations can be made regarding children's reading material. While we recognize that many other characteristics must be accounted for when studying readability compared to legibility, the findings herein are of significant import to studies of readability. After all, a word or passage cannot be read unless the letters are legible. This is of particular importance when considering a population of early readers. We recommend using the results of this study to further investigations on the readability of fonts for children.

REFERENCES

Anderson, P.V. (1987). *Technical writing: A reader-centered approach* (3rd ed.) pp. 378. United States of America: Harcourt Brace & Company.

Aslin, R. N. (1985). Oculomotor measures of visual development. In G. Gottlieb & N. A. Krasnegor (Eds.), *Measurement of audition and vision in the first year of postnatal life: A methodological overview* (pp. 391-417). Norwood, NJ: Albex.

Aslin, R. N. (1987). Visual and auditory development in infancy. In J. D. Osofsky (Ed.), *Handbook of infant development* (2nd ed., pp. 5-97). New York: John Wiley & Son's Inc.

Bachmann, B. A. (1991). Getting ready to read: Learning how print maps to speech. Washington, D.C.: National Institute of Child Health and Human Development, U.S. Department of Health and Human Services.

Bernard, M. L., Chaparro, B. S., Mills, M. M., & Halcomb, C. G. (2002). Examining children's reading performance and preference for different computer-displayed text. *Behaviour and Information Technology, 21,* 87-96.

Bouma, H. (1971). Visual recognition of isolated lower case letters. *Vision Research, 11,* 459-474.

Charness, N., & Dijkstra, K. (1999). Age, luminance, and print legibility in homes, offices, and public places. *Human Factors, 41(2),* 173-193.

Chauncey, C. (1986). The art of typography in the information age. *Technology Review,* Feb-Mar, p26+.

Erdmann, R. L., & Neal, A. S. (1968). Word legibility as a function of letter legibility, with word size, word familiarity, and resolution as parameters. *Journal of Applied Psychology, 52,* 403-409.

Gaines, R. (1969). The discriminability of form among young children. *Journal of Experimental Child Psychology, 8,* 418-431.

Gervaise, M. J., Harvey, L. O., & Roberts, J. O. (1984). Identification confusions among letters of the alphabet. *Journal of Experimental Psychology: Human Perception and Performance, 10(5),* 655-666.

Gould, J., Alfero, L., Finn, R., Haupt, B., et al. (1987). Reading from CRT displays can be as fast as reading from paper. *Human Factors*, 29(5), 497-517.

Hanna, P. R., Hodges, R. E., & Hanna, J. S. (1971). The alphabetic base of spelling. In Barbe, W. B., Francis, A. S., & Braun, L. A. (Eds.), *Spelling. Basic skills for effective communication* (pp. 150-164). Columbus, Ohio: Zaner-Bloser.

Hughes, L. E., & Wilkins, A. J. (2002). Reading at a distance: Implications for the design of text in children's big books. *British Journal of Educational Psychology, 72,* 213-226.

Jha, S. S., & Daftuar, C. N. (1981). Legibility of typefaces. *Journal of Psychological Researches, 25,* 108-110.

Kaye, D. B., Brown, S. W., Post, T. A., & Plude, D. (1981). The development of letter processing efficiency. *Memory & Cognition, 9(4),* 378-388.

Kowler, E., & Martins, A. J. (1982). Eye-movements of preschool children. *Science, 215,* 997-999.

Krueger, L. E. (1973). Effect of letter-pair frequency and orientation of speed of "same" – "different" judgments by children and adults. *Bulletin of the Psychonomic Society, 2(6B)*, 431-433.

Kruk, R. S. & Muter, P. (1984). Reading continuous text on television screens. *Human Factors*, 26 (3), 339-345.

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

Lannon, J. M. (Ed.). (2000). Chapter 15: designing pages and documents. *Technical Communication* (8th ed.). (pp. 304-322) New York: Addison Wesley Longman, Inc.

Legge, G. E., Pelli, D. G., Rubin, G. S., & Schleske, M. M. (1985). Psychophysics of reading: I. Normal vision. *Vision Research, 25(2),* 239-252.

Legge, G. E., Rubin, G. S., & Leubker, A. (1987). Psychophysics of reading: V. The role of contrast in normal vision. *Vision Research, 27(7),* 1165-1177.

Liberman, I.Y., Shankweiler, D., & Liberman, A. (1989). The alphabetic principle and learning to read. Washington, D.C.: National Institute of Child Health and Human Development, U. S. Department of Health and Human Services.

Mansfield, J. S., Legge, G. E., & Bane, M. C. (1996). Psychophysics of reading: XV. Font effects in normal and low vision. *Investigative Ophthalmology & Visual Science, 37,* 1492-1501.

Mills, C. B., & Weldon, L. J. (1987). Reading text from computer screens. *ACM Computing Surveys, 19(4),* 329-357.

Pinel, John P, J. (2000). *Biopsychology* (4th ed.). Boston: Allyn and Bacon.

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

Sanford, E. C. (1888). The relative legibility of the small letters. *American Journal of Psychology, 1,* 402-435.

Sanocki, T. (1988). Font regularity constraints on the process of letter recognition. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 472-480.

Slater, A. (1998). Perceptual development. *Visual, auditory, and speech perception in infancy.* East Sussex, UK: Psychology Press.

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

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

Tinker, M. A. (1959). Print for children's textbooks. *Education, 80,* 37-40.

Tinker, M. A. (1963). *The Legibility of Print.* Ames, Iowa: Iowa State University Press.

Yager, D., Aquilante, K., & Plass, R. (1998). Rapid communication: high and low luminance letters, acuity reserve, and font effects on reading speed. *Vision Research, 38,* 2527-2531.