

Effects of Three Typographical Variables on Speed of Reading

Richard H. Wiggins, Ph.D.

An examination of the literature on typographical research reveals a wide gap between research findings and practices. Further, there is considerable variation in the results reported by researchers, whose criteria for measurements and experimental methods have differed.

One of the earliest experiments on record was conducted in the 1790's by Anisson, the head of Imprimerie Nationale. Citizen Sobry reported in a speech in 1800 before the Société Libre des Sciences, Lettres et Arts of Paris that Anisson printed two one-page specimens, with one set in a modern Roman face of the Didot manner and the other composed in the Garamond style. The two pages were placed before experts who were asked to read the two specimens at various distances until the print was no longer distinguishable. Citizen Sobry stated the Garamond derivative "was readable several stages after Didot's characters had become indistinguishable."¹

The earliest two studies listed by Pyke in his historical résumé are T. C. Hansard in 1825, whose judgments were strictly *a priori*, and Charles H. Babbage, who in 1827 employed a majority vote to determine "facility in reading of different shades of paper."² R. L. Pyke listed no more studies until 1878 when Professor Emile Javal of the University of Paris began his pioneering studies. Javal made judgments on the basis of distances at which letters were distinguishable and also the

¹ "A Translation of the Reports of Berlier and Sobry on Types of Gille," Edited with an Introduction by Daniel Berkeley Updike, *The Fleuron*, No. VI (1928), 181.

² R. L. Pyke, *Report on the Legibility of Print* (London: His Majesty's Stationery Office, 1926), pp. 64-101.

visibility of letters under dim lighting. Another dimension was added to research on reading behavior in 1879 when Javal reported that the eye moves in quick, short movements during reading. Javal's initial investigations marked the beginning of hundreds of experiments exploring the problems of reading printed matter as psychologists, oculists, hygienists, physiologists, educators, lighting engineers, and others conducted studies.

From the standpoint of volume of investigations, the high point of typographical research was reached in the 1930's and early 1940's.³ Even though experiments during the period improved tremendously in sophistication as to criteria and statistical controls, criticisms leveled at the distance criteria used in the earlier period continued to be expressed. Frederic W. Goudy, the dean of American type designers, asked parenthetically concerning the Roethlein finding⁴ that New Gothic was the most "legible" for lower-case letters, "Ye Gods! and has it come to this?"⁵

Among the typographical variables that have concerned newspaper executives during the last two decades are line lengths (column widths), lower-case alphabet lengths, and the question of uneven versus even right margins.

The 13-pica line was regarded as the standard column width for newspapers up to World War I.⁶ Scarcity of newsprint and rising costs caused newspapers to reduce widths, and during the 1930's the 12-pica column became the standard. The American Newspaper Publishers Association reported that 96.8 per cent of its members were setting 12-pica lines when the news services inaugurated Teletypesetter tape transmission in 1951.⁷ By June, 1962, only 8 per cent of the ANPA papers included in an analysis were setting type 12 picas wide. On October 14, 1963, the Teletypewriter wire service transmission

³ John Scott Davenport and Stewart A. Smith, "Effects of Hyphenation, Justification and Type Size on Readability," *Journalism Quarterly*, XLII (summer, 1965), 382.

⁴ Barbara Elizabeth Roethlein, "The Relative Legibility of Different Faces of Printing Types," *American Journal of Psychology*, XXIV (1912), 1-36.

⁵ Goudy, *Typologia* (Berkeley and Los Angeles: University of California Press, 1940), p. 144.

⁶ John E. Allen, *Newspaper Designing* (New York: Harper & Brothers, 1947), p. 31.

⁷ *General Management Bulletin*, No. 31, June 27, 1962.

standardized on 11-pica widths when utilizing fonts with lower-case alphabet lengths of 118.1 points.

Although researchers had conducted experiments to determine the ideal line length for certain type sizes, none had studied specifically the effect of the reduction in width of newspaper columns on reading. Sir Cyril Burt stated that the 2-inch column was too narrow when the larger type sizes are used, and noted, “But the needs of the newspaper reader form an entirely different field of research, in which hardly any experimental work has as yet been attempted. . . .”⁸

Paterson and Tinker recommended line widths in the “neighborhood of 19 picas.”⁹ For 8-point solid, their specific recommendations were 13 to 24 picas.¹⁰

One rule of thumb often used to describe the ideal line length is that the line should contain between one and one-half alphabets and two alphabets, that is, 39 to 52 characters. Burt favored “fifty-five to eighty characters and spaces, ten to twelve words, and two to three lower-case alphabets.”¹¹

When the 11-pica TTS tape transmission was being planned, the use of a font with the longer lower-case alphabet length of 118.1 points (in comparison to the a-z length of 110.6 points) “so as to obtain increased legibility” was advocated by Charles H. Tingley, Managing Director of the ANPA Research Institute, Inc.¹²

One of the eight “qualities” Pyke measured in an experiment was the a-z length. He reported, “The coefficient for the a-z length contains less force than appears, because four of eight faces had the same length.”¹³ Burt observed, “In studies of legibility the influence of the width or ‘set’ of a type face has been almost wholly ignored. And, although we ourselves have made no direct investigation of this aspect, we believe that its importance is by no means negligible. . . .”¹⁴

A third variable affecting potential savings in typesetting is the

⁸ Burt, *A Psychological Study of Typography* (New York: Cambridge University Press, 1959), footnote 1, p. 17.

⁹ Donald G. Paterson and Miles A. Tinker, *How to Make Type Readable* (New York: Harper and Brothers, 1949), p. 156.

¹⁰ *Ibid.*, p. 149.

¹¹ Burt, p. 14.

¹² *Newsletter*, No. 37 (ANPA Research Institute, Inc.), August 1, 1962.

¹³ Pyke, p. 53.

¹⁴ Burt, footnote 1, p. 14.

question of uneven right margins versus the so-called justified right margins. The principal objection to uneven right margins is one of aesthetics. However, some typographers have felt that uneven space between words in order to produce justified lines slows down the reading process. The late Eric Gill, type designer and typographer, favored at least a compromise to provide more even spacing between words and at the same time have lines that were not “unpleasantly uneven.” He wrote, “But even spacing is of more importance typographically than equal line length. Even spacing is a great assistance to easy reading”¹⁵ Gill favored close spacing, “Provided that words are really distinct from one another, they should be set as close as possible.” A slightly different view was expressed by Burt, “For rapid reading a *moderately* close word-spacing is helpful.”¹⁶ Statements on spacing between words are largely subjective expressions, as researchers have neglected this variable.

Assumptions and Hypotheses

Several basic assumptions were made in the planning stage for this research. Even though the typographical effects were small, it was felt that with adequate controls these effects could be measured. This assumption was contrary to the one Pyke made when he stated, “The hypothesis is here put forward that extremely large typographical differences must be present before it is possible to say that there is any difference in the objective legibility of type.”¹⁷

Pyke had observed that “Physical obstacles prevent absolutely the attainment of some of the most obvious theoretical desiderata.”¹⁸ He had to use type faces that differed in qualities so that he was not manipulating only one variable. For example, his faces of the same a-z length varied in other qualities.

Type faces which varied only in lower-case alphabet length within one size have become available in recent years, so that it would be possible experimentally to measure this variable.

It was further assumed that perhaps some experimenters might not

¹⁵ Gill, *An Essay on Typography*, (London: J. M. Dent & Sons, Ltd., 1954) p. 88.

¹⁶ Burt, p. 12.

¹⁷ Pyke, p. 60.

¹⁸ *Ibid.*, p. 40.

have controlled or have been able to control “physical obstacles” to which Pyke referred.

Early research on legibility emphasized the visual aspects of seeing characters. It was assumed that the best measurement of the variables under study would be to use speed of reading which incorporated a comprehension check. Pyke asserted that “To read means to obtain meaning from written or printed symbols.”¹⁹

1) Since research has indicated that line length affects legibility, it was hypothesized that the reduction of newspaper column widths has significantly affected reading speed. It was further hypothesized that reading was slowed down by the reduction in column widths.

2) Since type fonts with different lower-case alphabet lengths affect the amount of material that can be printed within a given line length, it was hypothesized that differences in lower-case alphabet lengths affect reading speed significantly. It was further hypothesized that the fonts with the shortest lower-case alphabet lengths provided the fastest reading speed of printed text material.

3) It was hypothesized that the combination of the widest single newspaper column and the shortest lower-case alphabet length would be read the fastest.

4) It was hypothesized that constant spacing between words made the reading task easier even though the right margins were uneven, and that justified lines with uneven spacing between words made the reading more difficult.

5) It was hypothesized that the number of characters in a line of printing affects speed of reading, and it was planned to test experimentally the rule of thumb that the ideal line should contain approximately 39 to 52 characters.

Method and Procedure

The University of Minnesota Press granted permission to the experimenter to use “Tinker Speed of Reading Test, Form I and II,” by Dr. Miles A. Tinker. These two tests consist of statements totalling about 30 words each. Near the end of each is one incorrect word which the subject marks. Thus comprehension is necessary in order to mark this word correctly.

¹⁹ Ibid., p. 25.

Form I was set in 10-point Baskerville (Mergenthaler No. 10 delta 378) on a 10-point slug.

Form II contained the typographical variables. Nine specimens were prepared for the first experiment. Specimens were set in each of the three line lengths using three type fonts, Regal No. 1B TTS (font No. 2368 with lower-case alphabet length of 110.6 points), Regal No. 1 TTS (font No. 2345 with lower-case alphabet length of 118.1 points), and Regal No. 2 (font No. 1718 with lower-case alphabet length of 126 points). Regal is an Intertype Company face. Fonts No. 1B and No. 1 had x height of .054 inch, and font No. 2 had an x height of .057 inch.

The second experiment required 10 specimens, including the 10-pica specimen set in Regal No. 2 with even right margins for use in the first experiment. Regal No. 2 was used for all specimens of Form II in the second experiment. Two specimens, one with even right margins and the other with uneven right margins, were set for each line length.

The specimens which contained the variables were set on 8-point slugs. The Intertype Company considers .112 inch as the standard for 8 points. It was decided by the researcher that .115 inch would be the maximum thickness of slugs. When the slugs were checked by use of a micrometer it was found that most of the slugs exceeded the .115-inch maximum. Since leading is a variable that affects speed of reading, most of the slugs were processed through a milling machine in order to obtain a thickness of .115 inch or slightly less.

The five line lengths in the second experiment were selected because the Intertype Ready Reckoner, the copyfitting guide of the Intertype Company, gave the character counts for 8-point Regal No. 2 as 27 for 10 picas, 38 for 14 picas, 52 for 19 picas, 66 for 24 picas, and 79 for 29 picas. Thus these pica lengths theoretically would vary only one character from one alphabet length or 26 characters, one and one-half alphabets or 39 characters, 52 or two alphabets, 65 or two and one-half alphabets, and 78 or three alphabet lengths.

If this researcher had had available the required auxiliary units to a computer, a program could have been prepared to instruct the computer to make lines conform to a specified number of characters with variations to be made only when syllables required such fluctuations.

Form II was placed on punched cards which were processed through a computer to make a frequency distribution of the characters. By

calculating the space occupied by printed characters, the average space between words in full lines in the specimens set with even right margins was determined. Also, the average number of characters in full lines of type was obtained in order to compare with the figures given in the copyfitting guide.

The average space in the nine specimens for the first experiment varied from .064 inch (slightly less than $4 \frac{3}{4}$ points, which is .0657 inch) to .074 inch (about $5 \frac{1}{2}$ points, which is .0738 inch). The five specimens with variable spacing in the second experiment had average spaces that varied from .074 inch down to .056 inch. The specimens with constant spacing used a thin space that was .0311 inch wide (less than $2\text{-}1/4$ points, which is .0312 inch). Thus the specimens with even right margins in the second experiment had spaces that averaged a little more than twice the size of the thin space used in the specimens with uneven right margins.

The specimens in the first experiment contained one or two characters fewer than the estimates given in the copyfitting guide. The difference was as high as five fewer characters in one specimen in the second experiment. Copyfitting tables are based on the assumption that, on the average, 28.5 characters will equal the lower-case alphabet length of a type font. The actual length of 28.5 characters in the Form II specimens was calculated. The averages for 28.6 characters in the test forms was 3.5% to 4% longer than the lower-case alphabet length given for the type fonts.

When the actual length of 28.5 characters in the test form was substituted for the lower-case alphabet length of the type font, it was found that the specimen varied about one character from the theoretical estimate in the copyfitting table.

The number of lines ending with word divisions was determined. Hyphenation has been suggested as a variable affecting speed of reading. Davenport and Smith included hyphenation in a study of three variables, and found no significant effect for any of the three variables.²⁰

Experimental Design

A factorial design was selected for the two experiments so that two variables could be measured at the same time, and also any interaction between the two variables could be measured. The first experiment

²⁰ Davenport and Smith, pp. 382–388.

contained a line-length variable of 10, 11, and 12 picas and three lower-case alphabet lengths. The second experiment contained the spacing variable—constant spacing to give uneven right margins and variable spacing to produce even right margins—and the line-length variable of 10, 14, 19, 24, and 29 picas.

The statistic selected was analysis of variance or covariance. “Tinker Speed of Reading Test, Form I” provided an X measure which was the initial speed of reading abilities of the subjects.

The X score was used to rank subjects in descending order from the highest X score to the lowest X score. The students in each replication would be those with the smallest difference between the highest and lowest initial reading scores.

It was planned to make the initial calculations for analysis of variance since that statistic requires fewer calculations. It is desirable to have subjects within a replication as closely matched as possible in abilities when variance is used to measure treatments such as typographical variables in speed of reading.²¹

Should analysis of variance indicate that variations in abilities of the subjects required a statistical control to adjust for those differences, the X scores would be used to provide this additional precision in the computation of analysis of covariance. Also, a coefficient of regression

TABLE 1: *Analysis of Variance on 162 Observations for 3 Lower-Case Alphabet Lengths (Type Faces) and 3 Line Lengths*

<i>Source of Variation</i>	<i>Degrees of Freedom</i>	<i>Sums of Squares</i>	<i>Mean Squares</i>	<i>F</i>
Total	161	91,978		
R (Replications)	17	34,626	2,036.8	5.02***
A (Type Faces)	2	428	214.0	
B (Line Lengths)	2	1,508	754.0	1.86
AB (Type Faces x Line Lengths)	4	257	64.3	
Error	136	55,159	405.6	

*** $p < .001$

²¹ John E. Alman and David M. White, “Statistical Methods in Communication Research,” in Ralph O. Nafziger and David M. White, eds., *Introduction to Mass Communications Research*, Rev. Ed. (Baton Rouge: Louisiana State University Press, 1963), p. 161.

TABLE II: *Analysis of Covariance On 18 Samples For 3 Lower-Case Alphabet Lengths (Type Faces) and 3 Line Lengths*

<i>Source of Variation</i>	<i>DF</i>	<i>Sums of x²</i>	<i>Sums of xy</i>	<i>Sums of y²</i>	<i>df</i>	<i>Adjusted sums of y²</i>	<i>MS</i>	<i>F</i>
R	17	43,565	31,788	34,626				
A	2	96	192	428	2	151	75.5	
B	2	86	217	1,508	2	1,174	587.0	2.93
AB	4	213	41	257	4	523	131.0	
Error	136	31,512	29,776	55,159	135	27,023	200.0	

F of 2.93 for B (line lengths lacked .1 of a point of being significant to the .05 level).

could be obtained to adjust the Y means based on the Y scores made on the specimens which contained the typographical treatments.

The subjects read Forms I and II for 20 minutes each.

An F to the .05 level was established as the criterion for rejecting the null hypotheses of no differences in the effects of the three variables and interactions. The line-length variable was included in both experiments.

Analysis of variance computed on the first 162 observations had an interaction of less than one, the effects of the two variables were non-significant, and only replications yielded an F that was significant to the .001 level. There was a large error term. In order to improve the precision of the analysis it was decided to utilize the X-measure scores for analysis of covariance computations (see Table II).

With covariance, the data on the 162 observations yielded for line lengths an F of 2.93, which lacked .1 of a point of being significant to the

TABLE III: *Analysis of Covariance for 324 Observations for 3 Lower-Case Alphabet Lengths (Type Faces) and 3 Line Lengths*

<i>Source of Variation</i>	<i>DF</i>	<i>Sums of x²</i>	<i>Sums of xy</i>	<i>Sums of y²</i>	<i>df</i>	<i>Adjusted sums of y²</i>	<i>MS</i>	<i>F</i>
R	35	476,871	438,847	430,231				
A	2	445	1,424	4,621	2	2,280.5	1,140.3	3.95*
B	2	189	1,260	10,776	2	8,522.7	4,261.4	14.73***
AB	4	468	320	1,352	4	1,167.7	291.9	1.01
Error	280	70,687	67,593	145,344	279	80,709.6	289.3	

* p < .025 *** p < .001

.05 level. Covariance produced an adjusted mean squares for the error term that was 51.7% below the one for variance.

The experiment was enlarged to 324 observations. Covariance for the 36 samples produced for type faces an F of 3.95 that was significant to the .025 level, and the line length F was 14.74, significant to the .001 level (see Table III). The F for interaction was 1.01. Both of the null hypotheses of no difference in the means for the two variables being measured were rejected.

The coefficient of regression was computed to adjust the Y scores so as to examine the data with respect to speed of reading for the two variables. The type font with the shortest lower-case alphabet length was read 6.1 items or 3.8% faster than the font with the longest alphabet length. The 12-pica line was read 12.22 items faster or 7.7% in comparison to the 10-pica line.

The Y means for treatment combinations were adjusted and arranged according to characters in lines (see Table IV).

The rank order of the number of characters in specimens corresponded to the rank order of items read except for two instances. These two differences are attributable to chance fluctuations in those two cells and adjacent cells because observed effects, under the laws of chance, are more or less than the true effects of the treatment combinations. As the number of characters increased in the line for 10, 11, and 12 picas, items read also increased except for the two exceptions.

TABLE IV: *Treatment Combinations Arranged According to Characters Per Line With Adjusted Y Means*

<i>Treatment Combination</i>	<i>Average Number of Characters Per Line in Specimen</i>	<i>Estimate of Characters in Copyfitting Guides</i>	<i>Adjusted Y Means</i>
Regal No. 2, 10 picas	25.5	27	153.4
Regal No. 1, 10 picas	26.8	29	159.4
Regal No. 2, 11 picas	28.3	30	162.2
Regal No. 1B, 10 picas	28.5	31	158.9
Regal No. 1, 11 picas	30.0	32	163.1
Regal No. 2, 12 picas	30.8	33	169.2
Regal No. 1B, 11 picas	31.8	34	170.7
Regal No. 1, 12 picas	32.8	34	165.8
Regal No. 1B, 12 picas	34.5	37	173.6

TABLE V: *Analysis of Covariance for 300 Observations
On Spacing and Line Lengths*

<i>Source of Variation</i>	<i>DF</i>	<i>Sums of x²</i>	<i>Sums of xy</i>	<i>Sums of y²</i>	<i>df</i>	<i>Adjusted sums of y²</i>	<i>MS</i>	<i>F</i>
R	29	302,257	289,080	286,174				
A	1	44	144	466	1	235	235	
B	4	863	1,669	5,650	4	3,284	821	3.10*
AB	4	182	639	2,806	4	1,771	442.78	1.67
Error	261	117,469	111,459	149,953	170	44,194	260	

* $p < .025$

The subjects in the first experiment had an average age of 18.9 years. The most important population characteristic of the 324 individuals was their X-measure mean of 168 on the "Tinker Speed of Reading Test, Form I." Thus their hypothetical parent population had an initial speed of reading ability of 168 on this test.

Analysis of covariance on 300 observations in the second experiment yielded an F of less than one for spacing, and thus the null hypothesis for this variable was retained (see Table V). The F of 3.10 for line lengths was significant to the .025 level, and this null hypothesis was rejected. Interaction of F 1.67 was significant to the .20 level.

The adjusted Y means on the variables and treatment combinations are shown in Table VI. Even though the pica was actually the variable that was manipulated to vary the line length, hypothesis No. 5 hypothesized that the ideal line was approximately 39 to 52 characters.

The 14-pica line, which approximated the 39-character ideal as a minimum, was read faster than the 10-pica line. The 14-pica line was the optimum for constant spacing. The means for the constant spacing

TABLE VI: *Adjusted Y Means Arranged According to
Kind of Spacing One Line Length*

<i>Line Length in Picas</i>	<i>10</i>	<i>14</i>	<i>19</i>	<i>24</i>	<i>29</i>
Constant Spacing	163.3	176.9	171.4	172.0	172.3
Variable Spacing	165.7	168.8	175.8	165.6	170.1
Means for Both Kinds of Spacing*	164.6	172.9	173.7	168.8	171.7

* These means are computer calculations as adjusted by using the coefficient of regression of .9326.

specimens were lower for 19 picas than in comparison to the 24- and 29-pica specimens. The means for the 24-pica specimen with variable spacing was one-tenth of a point lower than the 165.7 means for 10-pica specimen with variable spacing.

The 300 subjects in the second experiment had an average age of 19.3 years. The key population characteristic of the subjects was the X-measure mean of 173.7 made on "Tinker Speed of Reading Test, Form I."

Discussion and Conclusions

The measurement of effects of independent variables manipulated in typographical specimens requires that adequate control be maintained to prevent other variables from being introduced into the experiments. Slug thicknesses should be measured in thousandths of an inch in order to obtain a more precise measurement to insure that leading is not added.

The initial computations on 18 samples in the first experiment in this study demonstrated the desirability for measuring the initial speed of reading of subjects. This provided a statistical control for computing analysis of covariance so as to compensate for the variation in reading abilities of the subjects. The data in the first experiment supported the hypothesis that speed of reading is decreased by the reduction of column widths from 12 picas down to the narrower widths of 11 and 10 picas.

The type font with the lower-case alphabet length of 110.6 points was read the fastest. Thus speed of reading was increased when newspapers changed from the longer lower-case alphabet length of 126 points down to 110.6 points, and data in this experiment indicated that speed of reading was decreased with the adoption of type fonts with lower-case alphabet lengths of 118.1 points in lieu of the 110.6 point length.

The first experiment statistically tested hypotheses dealing with the effects of variation on pica lengths and lower-case alphabet lengths on speed of reading. The combination of the two variables affected the number of characters in lines. The means for items read increased as the number of characters per line increased, except for two of the nine combinations of type fonts of lower-case alphabet lengths. The data on characters per line suggests that the number of characters that can be perceived in the eye span affects speed of reading.

The second experiment also yielded a very significant F on the line-length variable, but the spacing variable in the forms of variable spacing by use of space bands to give even right margins, and constant spacing by the use of thin spaces to provide uneven right margins was not significant. The specimens set with constant spacing contained more characters per line and thus contained fewer lines of type in comparison to the number of lines required for the specimen set with space bands. The specimens with constant spacing contained 31 per cent more lines ending with hyphens, compared to the word divisions in the specimens with even right margins. Some experimenters have considered hyphenation to be a variable affecting speed of reading. Thus the differences in hyphenation and the number of characters in lines probably produced the F of 1.67 ($p < .20$) for interaction in the second experiment. There was no interaction in the first experiment.

The adjusted means for both constant spacing and variable spacing supported the rule of thumb that 39 characters, or one and one-half alphabet lengths, provides faster reading than a shorter line such as 10 picas of 26 or fewer characters. However, the data for 19, 24, and 29 picas fluctuated so that the maximum line length in terms of characters and lower-case alphabet lengths remains in doubt on the basis of the second experiment.

It was concluded that experiments measuring the effects of even and uneven right margins require typographical specimens to be set line for line so as to prevent the introduction of other variables, such as hyphenation and differences in number of lines within the specimens. Thus the only variation should be the spacing between words.

Future experiments on the effects of spacing should include at least three different widths of spacing. Constant spacing with uneven right margins appears to be the best method for introducing three or more spacing widths, as even right margins will require that spacing be varied.

It is concluded that the rule of thumb based on the assumption that narrow and wide characters occur in a ratio that results in 28.5 characters, on the average, equalling the lower-case alphabet length of a font should be restudied by utilizing computers. This would result in more accurate copyfitting tables.

Many writers have stated that the ideal line length should be one which contains a certain number of syllables and words. So far, experimenters have had to select a line of a certain width in picas, and

after the type was set it could be determined how many characters the average full line contained. With the development of more sophisticated computer programs, future experimenters should be able to instruct the computer to produce a tape with a specified number of characters or syllables within a plus or minus range. Thus it would be possible to test experimentally the ideal number of syllables or words for a line length.

One of the three type fonts used in the experiment differed in x -height. This apparently had a slight effect on speed of reading, and the x -height variable should be studied experimentally in future experiments in typography.