Wide vs. Narrow Paragraphs: An Eye Tracking Analysis

David Beymer[1], Daniel M. Russell[1], and Peter Z. Orton[2]

[1] IBM Almaden Research Center, 650 Harry Road,
San Jose, California 95120 USA
[2] IBM On Demand Learning, 20 Old Post Road,
Armonk, New York 10504 USA
{beymer, daniel2, porton}@us.ibm.com

Abstract. How wide should paragraphs be formatted for optimal reader retention and ease of reading? While everyone is familiar with the narrow, multi-column formatting in newspapers and magazines, research on the issue is not consistent. Early work using printed media favored narrow formatting, while more recent work using computer monitors has favored wider formatting. In this paper, we approach this issue by using eye tracking analysis of users reading material on instructional web pages. In our experimental system, subjects read the material using an instrumented browser that records all HTML content and browser actions, and their eye gaze is recorded using a nonobtrusive, "remote" eye tracker. Comparing the wide and narrow formatting conditions, our analysis shows that for narrow formatting, subjects (a) read slightly faster, (b) have fewer regressions, (c) retain more information in a post-test of the material, but (d) tend to abandon the ends of longer paragraphs.

1 Introduction

In this paper, we address the question: how is reading behavior affected by narrow vs. wide paragraph formatting? Are there observable differences between readers given the same material, but differing paragraph widths? Early work by typographers, psychologists, and designers focused on printed material. Using reading speed as a metric, these experiments favored shorter line length (Tinker [6] favored 80 mm), which is recognizable today in the narrow columns typically seen in newspapers and magazines. The intuition is that the "return sweep" from the end of one line to the beginning of the next is more difficult in wide paragraphs because of the longer eye movement required by the reader. More recently, however, researchers in ergonomics and human factors have re-visited the issue with reading from computer screens [1,3,4,5,14]. In contrast with print studies, reading speed on computer screens favored longer lines (e.g. most of the monitor width) or at least showed no preference for shorter lines. To help reconcile these differences, it has been noted that monitors are typically further from the reader than hardcopy pages, so a narrow printed paragraph may subtend a similar visual angle on the eye as a wider paragraph on a monitor.

Eye gaze tracking provides a valuable tool for objectively measuring reading behavior. In eye gaze tracking systems, a camera records the eye gaze of a user, which in the case of reading is mapped to the words and lines of text being read. Amazingly, eye tracking technology has a history that stretches back 100 years, with early work coupling photographic recording methods with a "corneal reflection" technique that bounced light beams bounced off the subjects' eyes. Indeed, early application of this technology was used to study reading behavior. While readers intuitively may think that their eye gaze follows a continuous left-to-right motion, eye tracking studies showed that eye motion advances in discrete chunks across the page. A reader's eyes will actually stop, or fixate, on a set of characters for about 250 ms. This *fixation* is followed by a *saccade*, an eye movement of about 10 characters to the right, where the eyes will stop at the next fixation. A *regression*, or backwards eye movement in the text, is a sign that the reader is having difficulty understanding the material.

Given the tool of eye gaze tracking, what does it tell us about reading performance? Eye tracking has confirmed the intuitive notion, for instance, that the difficulty of reading material has an effect on reading. For example, material from a physics or mathematics textbook will cause longer fixations and more regressions than material from light fiction or newspapers [7]. Eye gaze tracking has also been used to study the effects of typography on reading [8]. In 1940, Patterson and Tinker [9] studied eye movements when subjects were presented with text formatted with different line lengths. From eye gaze, they measured fixation frequency and perception time (the sum of time spent in fixations), measures that have been found to correlate with reading speed. These measures favored an intermediate line length of just over 3 inches. In a longer line length tested (7 inches), the slowness of subjects was found to be explained by increased regressions and difficulty in return sweeps. These tests were done with printed text, not text presented on computer screens.

Eye tracking systems have evolved greatly since the photographic techniques used by Paterson and Tinker. While the corneal reflection method is still used today, the light beams now use nonobstrusive infrared light and recordings are measured using video techniques. A number of commercial eye gaze trackers exist, with some software systems performing reading analysis on recorded eye gaze.

In this paper, we revisit the question of how typography affects reading, focusing on the issue of line length. Relative to the 1940 study by Tinker and Paterson, we think that (a) the maturation of eye gaze tracking technology makes the recording less obtrusive and the analysis more automatic and less error-prone, and (b) given the earlier disagreement between print-based and computer-based line length studies, a study of online reading from monitors may provide a different result.

We have developed a tool, WebGazeAnalyzer (WGA) [10], for recording and analyzing the eye gaze of a user during a web browsing session. A nonobtrusive camera mounted inside the computer monitor bezel observes the eye gaze of the computer user, estimating where on the screen the user is looking. Reading is detected in the eye gaze by looking for the characteristic horizontal pattern of fixations and saccades. By intersecting the eye gaze location with the location of words on the web pages, we can tell what the user is reading, the reading speed, what they are re-reading, skipping, etc. Our recording tool records multiple data streams using a special instrumented version of Internet Explorer. We record: a movie of the computer screen, all

URLs visited by the user, a parsed version of the HTML content, and all Windows events such as scrolling and mouse clicks. After recording the data, we can browse it as a movie using a playback tool, or we can batch analyze a number of sequences to analyze reading behavior across different experimental conditions. In this paper, we use WebGazeAnalyzer to study reading behavior under wide and narrow formatted paragraphs.

2   Measuring Reading Using Eye Tracking

Given that WGA's recording system has recorded the eye gaze and web content from a particular experiment, how does WGA analyze the eye gaze to report reading statistics? While the mechanics of the recording system and reading analysis were largely reported in [10], we expand upon some of the details of how time, distance, and speed are measured, since those are the important statistics reported in this paper.

2.1   Low-Level Gaze Processing

From the raw eye gaze, our analysis system first detects fixations in the gaze stream using a dispersion-based approach [13]. Because of the 0.5 degree of error in the gaze estimate from our eye tracker, assigning fixations to URL text for reading analysis may be ambiguous – given the uncertainty region around a fixation, the fixation may potentially map to multiple lines of text. As detailed in [10], we address this uncertainty with a robust, line-based matching algorithm that assigns fixations to lines of text from the recorded web page. After this matching process, the system can map each fixation to a single line of text, allowing reading analysis to proceed.

    After matching, the analysis system next parses the fixation data into *forward reads*. A forward read is a grouping of consecutive fixations moving forward through the text with typical reading saccades (see Fig. 1). When processing the fixations to form forward reads, a forward read will be stopped when (1) a regression is encountered, (2) a forward saccade is too large and likely a forward skip through the text, or (3) the eye gaze moves to another line of text. Intuitively, forward reads are designed to capture when and where subjects are actively reading the text in detail. As shown in Fig. 1, the ith forward read detected, fr_i, covers a distance d_i that is padded to the left and right by the perceptual span [7]. Furthermore, let time t_i be the time spent in the forward read, defined as the time difference between the beginning of the first fixation and the end of the last fixation.
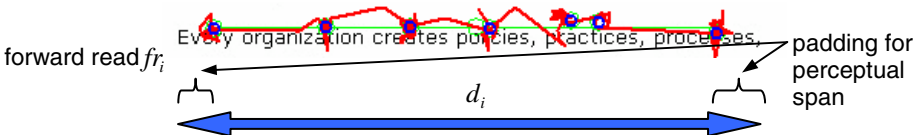
Fig. 1. For forward reads, the reading distance is padded by the perceptual span. Detected fixations are shown with small circles, interconnected by the raw gaze.

Typically in an experiment, the researcher is only interested in reading behavior for particular content-of-interest, whether it be specific sentences or entire paragraphs from the URL. While WGA offers flexibility in specifying content-of-interest down to the word level, in this paper we will analyze reading statistics over paragraphs. We now explain how forward reads are processed over the specified paragraphs-of-interest.

2.2 Instantaneous Measures

Given the short amount of time covered by forward reads, typically 1 to 2 seconds, measuring reading using forward reads alone gives us a raw, instantaneous measure of reading. Given a line of text from a paragraph-of-interest, the gaze matcher will assign some number of forward reads fr_0 … fr_{n-1} to the line. We measure the time

$$T1 = \sum_{i=0}^{n-1} t_i \tag{1}$$

as the total time the subject's eyes are actively reading forward through the text. Note that $T1$ does not count time spent in regressions, return sweeps, or distractions like scrolling or scanning. The text distance covered by these forward reads

$$D1 = \sum_{i=0}^{n-1} d_i \tag{2}$$

will include re-reads of the same material.

Our instantaneous speed measure $D1/T1$ is a low-level measure of eye velocity while reading. Given the low-level nature of this speed measure, it may be most useful as an input to some modeling process as opposed to a final reporting statistic.

2.3 Overall Measures

In measuring the overall speed of reading, we want to include the effects of regressions and return sweeps. Also re-reading the same material should count against the reader, as this slows down his or her overall performance. To account for this, we can re-use the forward reads as follows. For time, measure the **elapsed time** between and including the first and last forward reads

$$T2 = \text{(end time of } fr_{n-1}) - \text{(begin time of } fr_0) . \tag{3}$$

For distance, we only measure the first time material is read by projecting the forward reads onto the text lines and ignoring overlapping segments—we call this the reading coverage of the text lines. Given that the content-of-interest is composed of a number of text lines, we compute

$$D2 = \sum_{\text{all text lines}} \text{text line coverage} \tag{4}$$

which measures the amount of text read at least once (but does not double count re-reads).

Finally, the overall speed is computed as *D2/T2*. If the subject re-reads the same material in multiple passes separated by non-reading activities, each pass is treated separately with its own *T2* measurements, and speed is measured by $D2 / \Sigma\, T2$.

3 Experiment

Because our experimental system is designed to record and analyze web browsing sessions, we developed web pages with the same text content, but different paragraph widths. The two conditions used in our experiment include:

Cond. A: *Wide paragraphs*. Paragraphs width was 80% of screen width, measuring 9 inches on our monitor.

Cond. B: *Narrow paragraphs*. Paragraph width was 40% of screen width, measuring 4.5 inches.

The font used was Verdana 10 pt, and subjects are positioned approximately 60-70 cm from the monitor. In both conditions, scrolling is required to read the entire page. The experimental web page describes the different stages of "culture shock" of employees on overseas assignments, and it is a real page from IBM training course material. Fig. 2 shows an example of one paragraph in each of the two conditions.

Subjects were volunteers from a group of new managers attending management training school, so the web material was relevant to their background, motivations and interests; each subject had extensive practice (on the order of tens of hours) with these course materials before our test. There were 8 subjects in condition A (6 male, 2 female) and 8 subjects (1 male, 7 female) in condition B.

Before seeing the culture shock page, each subject read through 2 practice pages, and had to answer several basic data collection questions (e.g., native language, handedness, etc.) Subjects were told there would be a post study test of their retention of the content and that they should study the materials in anticipation of the test just as they had during their normal course of instruction.
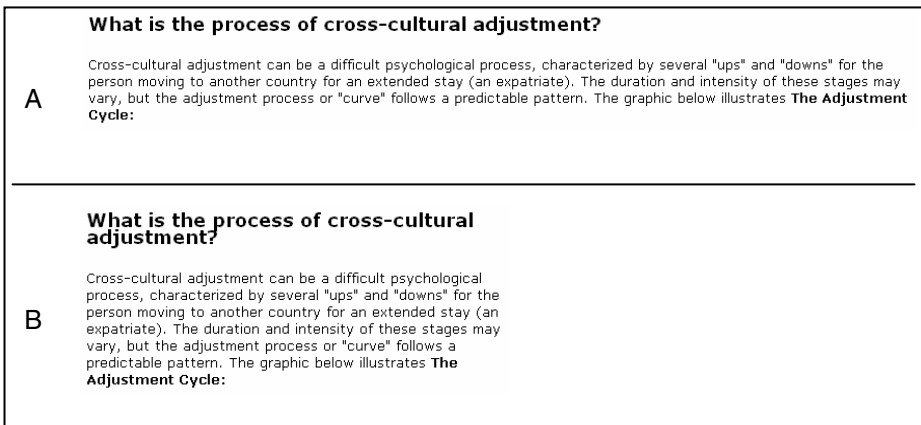
Fig. 2. In condition A, paragraph formatting is wide (9 inches on our monitor), while in condition B, paragraphs are narrow (4.5 inches on our monitor)

Eye tracking was done with the Tobii 1750 eye tracker [11] and two IBM T41 laptop computers. One laptop, the "gaze server", was devoted to running the Tobii eye tracking software. The second laptop, the "user machine", was the machine used by our subjects. On it ran our instrumented Internet Explorer browser, which records all URLs visited, HTML content (through the Document Object Model), and dynamic events such as scrolling [10]. We also record an event-driven movie recording of the user's screen through an adapted version of the VNC remote desktop system (see [12] for a related screen capture system).

4   Results

Reading analysis focuses on the first 5 paragraphs of the culture shock URL, looking for reading differences between the wide and narrow conditions. As mentioned in section 2, reading analysis proceeds by detecting fixations, matching them to lines of URL text, and finally grouping them into forward reads, which are the characteristic reading pattern of fixations and saccades in the eye gaze stream. Thus, for each paragraph, we can measure reading speed, regressions, the fraction of material read (vs. skipped), and return sweeps from one line of text to the next.

Comparing conditions A and B (the wide and narrow conditions, respectively), our reading analysis does show a slight speed advantage for B, but this is not caused by A being slowed down by return sweeps. One surprising result is that narrow paragraphs encourage the readers to skip material.

4.1   Reading Time and Coverage

Subjects in condition A (wide) spend more total time reading and read more of the material than B (narrow). For the first five paragraphs in conditions A and B, we measured reading time using the *T1* and *T2* metrics measured defined in section 2.

Subjects in condition A spend more time in both measures (*T1*: 13.22 sec, *T2*: 15.80 sec) than subjects in condition B (*T1*: 9.28 sec, *T2*: 11.35 sec). The extra time condition A subjects spend reading is used to read a greater fraction of the material, as compared to condition B. Fig. 3 shows time measure *T1* (top) and the fraction of paragraph material that is read (bottom). Note that those in condition B progressively read less and less of the paragraphs as they move down the page. This paragraph abandonment may be due to the longer paragraph height as a result of the reduced width. Perhaps this increased height leads to a perception of lengthy material and encourages skipping ahead. The difference between A and B in fraction read for paragraph 5 is statistically significant, $F(1, 14) = 4.655$, $p = 0.048$; for paragraphs 1-4, the differences in fraction read are not significant, but the trend seems clear.

This relationship between line length and reading coverage has not, to our knowledge, been examined before. This is primarily due to the difference in tasks given to our subjects versus, for instance, Paterson and Tinker [9]. Our subjects are given an open reading task, being told to freely read a web page. This gives them the
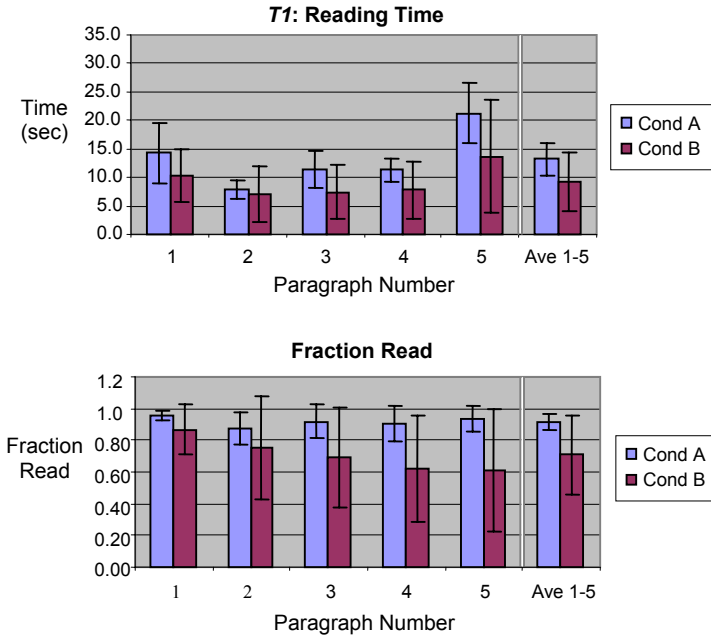
Fig. 3. Subjects spend more time reading the text under condition A (top chart). The fraction of paragraphs read remains high for condition A, but deteriorates for condition B (bottom). In this bar graph and subsequent ones, the error bars indicate the standard deviation in the distribution.

opportunity to skip around in the material, perhaps deciding to ignore some material. In [9], subjects are essentially forced to read all the material by the nature of the test used – the "Chapman-Cook Speed of Reading Test". In that test, subjects read a paragraph where a single word "spoils" the meaning, and the subjects are asked to identify the word. Thus, any influences of typography on *whether* to read the text are eliminated from [9].

4.2 Reading Speed

Subjects in condition B are slightly faster than those in condition A. Referring back to section 2, we measure speed in two different ways, an instantaneous measure of eye reading velocity and an overall measure of reading speed.

Subjects in condition B read faster than condition A under both metrics (see Table 1). As the table shows, condition B is faster by about 7% for instantaneous speed and 15% for overall speed, but, unfortunately, even the latter difference is not statistically significant ($F_{(1, 14)} = 1.035$, $p = 0.32$). The instantaneous speed result is in rough agreement with [9], where they saw a 7.8% difference in fixation frequency favoring short line lengths over long (fixation frequency is correlated to speed).

Table 1. Subjects in condition B read slightly faster than condition A (see text for details)

Speed	Instantaneous (pix/sec)	Overall (pix/sec)
Cond. A	275, $\sigma = 37$	161, $\sigma = 35$
Cond. B	295, $\sigma = 68$	185, $\sigma = 52$
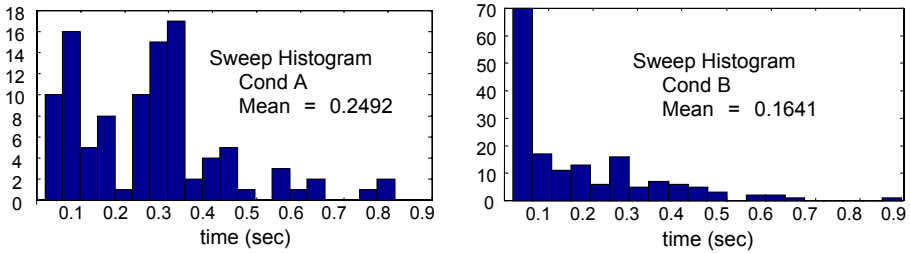
Fig. 4. Histograms of return sweep times for conditions A and B. The peaks near 0.1 sec are single saccades, while the peak in condition A at 0.3 sec is 2 saccades and a fixation.
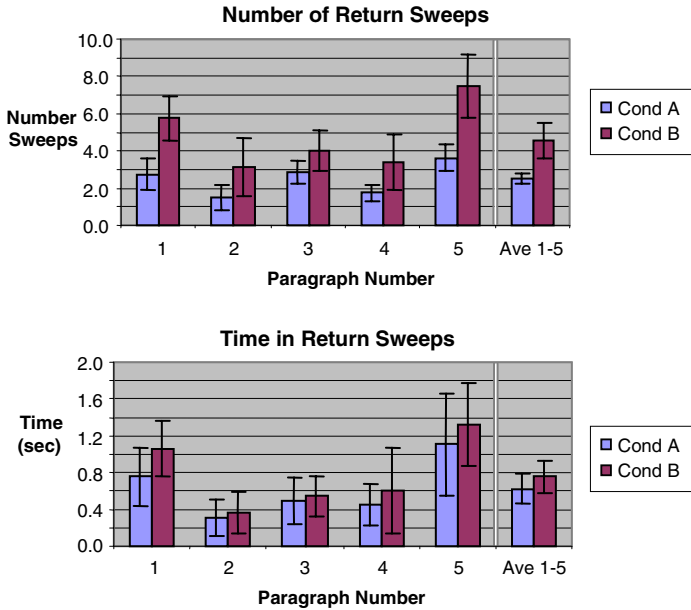
4.3   Return Sweeps

The long return sweeps in condition A take more time than condition B and often require an additional positioning fixation. As shown in Fig. 4, the histogram on return sweep times is bimodal for condition A and unimodal for condition B. For the short distance covered by return sweeps for condition B, the eye makes a single saccade from the end of the previous line to the beginning of the next line (see Fig. 5). Similarly for condition A, sometimes the eye performs the return sweep in a single saccade, yielding the first histogram peak. But quite often the eye requires an additional fixation and saccade to position the eye for reading the next line (Fig. 5). This additional fixation is always close to the beginning of the next line, and usually undershoots the return. This type of return sweep yields the second histogram peak in
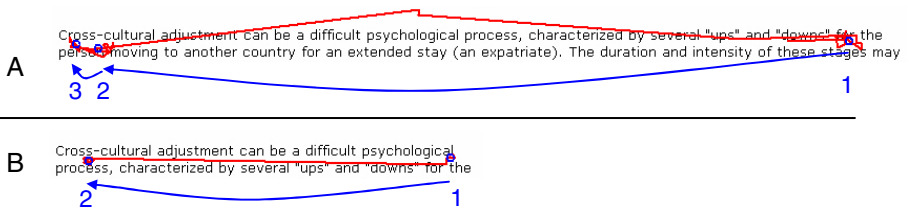
Fig. 5. In condition A, the long return sweep often requires an additional fixation (fixation 2 above) to position the eye for reading the next line. In condition B, the return sweep is most often performed with a single saccade. Detected fixations are shown with small circles, interconnected by the raw gaze.

Number of Return Sweeps

Time in Return Sweeps

Fig. 6. Condition B has more return sweeps than A (top chart), and requires more total processing time from the subjects (bottom)

Fig. 4 for condition A. (Bouma [2] also noticed "correction saccades" after "line saccades" in his eye tracking studies.) For the first five paragraphs in the test web page, the average sweep times are 0.25 sec for condition A and 0.16 sec for condition B.

4.4 Time in Return Sweeps

For condition B, the advantage of shorter return sweeps is outweighed by their increased number. While return sweeps require less time in condition B, the shorter paragraph formatting creates more of them for the reader to process. In Fig. 6 (top), we show the number of return sweeps detected for the first five paragraphs in the text; the numbers correlate well with the number of lines in each paragraph. Fig. 6 (bottom) shows the average time in return sweeps for conditions A and B, which is basically the product of the number of sweeps with the mean sweep time. Because subjects in condition A spend less time in return sweeps but have a lower speed, they must have some other speed handicap. This turns out to be increased regressions.

4.5 Regressions

As shown in Fig. 7, the subjects in condition A have a higher rate of regressions. Averaged over the first five paragraphs, the regression rate under condition A is 0.54 reg/sec, while it is only 0.39 reg/sec for condition B. The regression rate is computed by dividing the total number of regressions by *T2*, the total elapsed time. Regressions are used as a cue for reading difficulty – more difficult reading material will generate

more regressions than less difficult material. In this case, however, because the material is identical across conditions, the difference in regression rate must be from the increased line length. We hypothesize that the increased line length causes the eyes to occasionally make tracking mistakes, which requires backing up and re-reading. The increase in regressions for the longer line length condition is in agreement with Paterson and Tinker [9].
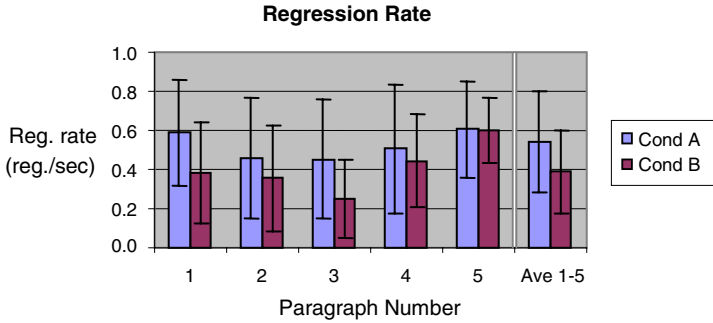

Fig. 7. Subjects in condition A have a higher regression rate than condition B

4.6 Retention

Subjects in condition B have better retention than condition A on a post test of the URL material. After reading the web page on culture shock, the subjects are given a 3-question multiple-choice test on the material. On average, subjects in condition A get 1.25 questions correct, while those in condition B score an average of 1.75 correct. This difference is even more interesting in light of the fact that condition A subjects spend more total time reading the material than condition B. Given this observation, another relevant statistic is the score divided by elapsed time with the material – getting a high score using less time is ideal. After normalizing each subject's score by $T2$, the elapsed time, we get an average normalized score of 0.081 points/sec for condition A and 0.191 points/sec for condition B. Using this time-normalized metric, there is a significant difference between the two conditions, $F(1, 14) = 6.798$, $p = .0207$.

5 Discussion

The experiment described here presents some basic tradeoffs in choosing paragraph width. Making the paragraph width wider reduces the number of return sweeps required for reading, but increases regressions and decreases retention. While the experiment presented here used two specific paragraph widths for testing, paragraph width can be finely tuned. It would be interesting to repeat the histogram analysis of return sweeps for a variety of widths to see when the auxiliary "positioning" fixation begins to appear as a function of width. This would suggest choosing a width just below this threshold.

Relative to the earlier work of Paterson and Tinker [9], our work is in agreement on (a) reading speed favoring shorter lines, and (b) increased regressions being recorded for longer lines. Moving beyond the results in [9], we showed that shorter line length tends to cause paragraph abandonment and hence less reading coverage. Also, a post-test of retention showed that the subjects performed better with shorter line length. Interestingly, with regards to reading speed, our results were more in alignment with the earlier work with print media as opposed to more recent work with online reading from monitors.

An interesting question arises as to why readers of the narrow widths did not continue reading as far to the end of the material as the readers of the longer widths. As noted, one possibility may be perception of a greater amount of material to be read in the tall columns. Another possibility may be that narrow-width text is associated with newspaper columns, which are almost universally written in the "inverted triangle" style. That is, newspaper stories put the most important information at the top of their column inches and become progressively less important as the column grows longer. Perhaps readers familiar with newspaper-style writing associated the narrow column widths to be written the same way, and thus concluded that the material near the end of the column was far less important to read. From a design perspective, important information should not be put in the second half of narrow-width paragraphs.

But perhaps the most important implication for message senders is that the shorter line lengths produced significantly greater retention. Instructional designers ought certainly to know that certain configurations of text provide a greater opportunity for material to be remembered, because learning is directly correlated with retention. The importance of this finding alone warrants further investigation of the implications of wide versus narrow column length web text.

References

1. Bernard, M., Fernandez, M., and Hull, S., The Effects of Line Length on Children and Adults' Online Reading Performance. *Usability News*, 4:2 (2002).
2. Bouma, H. Visual Reading Processes and the Quality of Text Displays. In *Ergonomic Aspects of Visual Display Terminals*, E. Grandjean and E. Vigliani, Eds. (1980), 101-114.
3. Duchnicky, R.L. and Kolers, P.A. Readability of Text Scrolled on Visual Display Terminals as a Function of Window Size. *Human Factors*, 25:6 (1983), 683-692.
4. Dyson, M.C. and Kipping, G.J. The Effects of Line Length and Method of Movement on Patterns of Reading from Screen. *Visible Language*, 32:2 (1998), 150-181.
5. McMullin, J., Varnhagen, C., Heng, P., and Apedoe, X. Effects of Surrounding Information and Line Length on Text Comprehension from the Web. *Canadian Journal of Learning and Technology*, 28:1 (2002).
6. Tinker, M.A. and Patterson, D.G. Studies of Typographical Factors Influencing Speed of Reading: Length of Line. *Journal of Applied Psychology*, 13:3 (1929), 205-219.
7. Rayner, K. and Pollatsek, A. The Psychology of Reading. Lawrence Erlbaum Associates, Hillsdale, NJ (1989)
8. Tinker, M. Legibility of Print. Iowa State University Press, Ames, Iowa (1963)
9. Paterson, D. and Tinker, M. Influence of Line Width on Eye Movements, *Journal of Experimental Psychology*, vol 27, pp. 572-577 (1940)

10. Beymer, D. and Russell, D. WebGazeAnalyzer: A System for Capturing and Analyzing Web Reading Behavior Using Eye Gaze. In *CHI 2005 Extended Abstracts*, ACM Press, pp. 1913-1916 (2005)
11. Tobii 1750 Eye-Tracker, Tobii Technology, http://www.tobii.se, April, 2005.
12. Li, S.F., Spiteri, M.S., Bates, J., and Hopper, A., Capturing and Indexing Computer-based Activities with Virtual Network Computing. ACM Symposium on Applied Computing, 601-603 (2000).
13. Salvucci, D.D. and Goldberg, J.H. Identifying Fixations and Saccades in Eye-Tracking Protocols. *Proc. of the Symposium on Eye Tracking Research & Applications* (ETRA), 2000.
14. Dyson, M. and Haselgrove, M. The Influence of Reading Speed and Line Length on the Effectiveness of Reading from Screen. *Int. J. Human-Computer Studies*, vol 54, pp. 585-612 (2001)