

How physical text layout affects reading from screen

MARY C. DYSON

Department of Typography & Graphic Communication, The University of Reading, 2 Earley Gate, Whiteknights, Reading RG6 6AU, UK; e-mail: M.C.Dyson@reading.ac.uk

Abstract. The primary objective of this paper is to critically evaluate empirical research on some variables relating to the configuration of text on screen to consolidate our current knowledge in these areas. The text layout variables are line length, columns, window size and interlinear spacing, with an emphasis on line length due to the larger number of studies related to this variable. Methodological issues arising from individual studies and from comparisons among studies are identified. A synthesis of results is offered which provides alternative interpretations of some findings and identifies the number of characters per line as the critical variable in looking at line length. Further studies are needed to explore the interactions between characters per line and eye movements, scrolling movements, reading patterns and familiarity with formats.

1. Introduction

1.1. Context

The rapid development of the World Wide Web (WWW), alongside our increasing use of word processors for creating documents, has increased the volume of material that we read from screen. As with reading printed material, there are many different ways of reading electronic texts, which may be determined by the reader's purpose, skills, or circumstances of reading. Web pages are frequently skimmed, rather than read in detail (Horton *et al.* 1996), but there are also circumstances where we read at a normal rate. If we can identify some ways of presenting text on screen which facilitate more effective reading, this may ultimately reduce the amount of documents we print from screen.

Extensive research has been conducted into the legibility of print, examining a range of typographical variables summarised in, for example, Tinker (1963, 1965), Zachrisson (1965), and Spencer (1968). Hartley

(1994) has moved this work on by developing guidelines for designing instructional text, combining current practice and more recent research findings. This approach addresses one of the criticisms that can be made of some early legibility research, which is that the studies lack internal validity due to the inappropriate choice of typographic material (Lund 1999). This tends not to be the case when there is contact between researcher and those with typographic knowledge, as discussed by for example Dyson (1999) and Lund (1999). This issue is discussed in more detail in the section on relationships between typographic variables.

Much of the detailed literature on typographical issues related to text presentation remains unreplicated on screens, as Dillon (1992) points out in his excellent review of empirical literature on reading from paper vs. screens. The situation was not very different four years later, as Muter (1996) observed that we do not know how to optimise reading from screen in relation to the layout of text. Even though there may be a range of layouts which are suited to different ways of reading, we need to start by identifying some of those variables which may influence reading effectiveness and those which may be less important.

This may seem to be a relatively unambitious aim, but the empirical research literature on screen reading is relatively immature. There is limited evidence of the effects of layout on different methods of reading but some work has looked at this, for example by speeding up reading (Dyson and Haselgrove 2001). A possible reason for the rather restricted set of studies to date may have its basis in suggestions that we can translate our knowledge of designing for print to screen (e.g. Kahn and Lenk 1993, Borchers *et al.* 1996). However, there have also been warnings against this. Kolars *et al.* (1981: 526–527) concluded that classical sources on the

printed page 'cannot be extrapolated wholesale to this newer medium'. These authors identify Huey (1908) and Tinker (1963) as the classical sources who made the initial exploration of factors affecting the reading of print. More specifically, Grabinger and Amedeo (1988) warned that standards derived from the classical sources were frequently used for screen text but their application to this medium had yet to be established.

In this paper, references are made to the legibility of print research only as an aid to interpreting results from reading from screen. As with most comparisons of the outcomes of different studies, there may be many factors which vary among studies and may therefore account for differences. These might include the measures, materials, and tasks used and participants' characteristics. Comparing print and screen adds another factor which, as discussed above, should be taken into account. Therefore, identifying divergent findings from studies of print and screen can contribute to the evaluation of research into reading from screen, but rarely can such differences be wholly attributed to the medium.

This paper aims to assess and interpret the current state of knowledge of empirical research on variables contributing to physical text format on screen. Due to the relative paucity of empirical studies specifically addressing these variables on screen in contrast to print, the paper is not intended as an extensive literature review.

1.2. *Scope*

The dates of previously published reviews on reading from screen (Mills and Weldon 1987, Dillon 1992, Muter 1996) generally limited them to citing experiments using display technology of the 1980s. These results may have questionable validity in relation to the higher quality displays which emerge in the literature on research into reading from screen published after about 1990. However, they have been retained in this paper to enable cautious comparisons with more recent findings. They also augment the rather limited number of studies in the area.

The reviews cited above cover comparisons between paper and screen, which appear to have been the starting point for research into reading from screen. As such comparisons have limited use in identifying what may facilitate more effective reading on screen, they are not included here unless they include investigation of typographic variables relevant to this paper. One of the difficulties with many of these studies is the use of different typography for text printed on paper and displayed on screen. This can be addressed by trying to equate the two conditions (e.g. Osborne and Holton

1988), or by choosing to optimise the format for the medium (e.g. Muter and Maurutto 1991).

Typographic variables included in this paper relate to configurations of texts, or text formats, which incorporate line length, columns, window size and interlinear spacing. Other variables, such as justified vs. unjustified text, treatment of paragraphs (indentation vs. additional interlinear space), and arrangement of headings have not been studied individually. However, some are included within studies looking at a combination of screen variables. As format can cover such a broad range of variables, e.g. colour of text and background, arrangement of white space, use of graphic devices (Grabinger and Osman-Jouchoux 1996), the scope of this paper has been limited.

Twyman (1982) describes configuration variables as extrinsic and distinguishes them from intrinsic (character) features. The only intrinsic feature that is considered is type size as it relates to line length, due to inter-relationships between variables. Experiments comparing different typefaces are not included in this paper, as this intrinsic variable falls outside the scope as defined above. Typeface may be considered an important typographic variable, sometimes regarded as synonymous with typography. However, the empirical literature does not reflect this importance, with a limited number of published studies on typefaces on screen, and no clear evidence of differences in their legibility.

The majority of studies described investigate the effects of the selected variables on reading rate, comprehension, eye movements and preference.

1.3. *Relationships between typographic variables*

Many typographers would maintain that typographical variables cannot be considered individually. Designers should make decisions on each variable in relation to other variables. There is some support for this from studying the visual processes of reading. Long line lengths are said to need more interlinear spacing to ensure that the eyes locate the next line down accurately when executing a return sweep towards the end of a line. The angle of the return sweep should not be too small (Bouma 1980).

Lund (1999) agrees that typographical variables should not be treated individually and formulates his argument in terms of the internal validity of the study (as mentioned earlier). He explains that keeping invariant any variables other than the target variable will result in differences in the ratios between variables, producing a confound. For example, as line length increases without a corresponding increase in interlinear spacing, the ratio between the two variables changes.

Following Lund's arguments, any differences attributed to longer lines might be due to a reduction in the relative amount of interlinear spacing. The solution proposed is to adjust variables to maintain constant relationships between them.

Underlying Lund's internal validity is the importance of reflecting the practices employed in good typographic design. However, this approach can be interpreted as replacing one confounding variable with another, depending on whether variables are defined as a single dimension or a higher order relationship. Also, the inter-relationships may be extremely complex.

This approach can be taken even further by combining the 'best' or recommended settings for a number of variables and comparing these layouts with a control condition of some description (Duin 1988, Muter and Maurutto 1991, de Bruijn *et al.* 1992, Grabinger, 1993). Combining variables in this way has been described as a 'kitchen sink' approach (Muter 1996). This approach requires some expertise in designing the test material to validate the selection of the levels of variables, taking into account the inter-relationships referred to above.

The alternative approach is to start by identifying the effect of individual typographic variables and then to test whether the effect of one variable depends on another using systematic replications. Muter (1996) discusses the problems that such interactions create and suggests that it is usually impractical to perform the huge number of experiments required. Rather surprisingly, Paterson and Tinker did study type size, line length and interlinear spacing of print with simultaneous and systematic variation of all three, summarised in Tinker (1965). However, such enormous studies have not been replicated for reading from screen. Instead, the range of possibilities has normally been constrained. Less expertise is necessary when selecting the level of a variable that will remain constant in an experiment. However, to ensure that the results will relate to normal reading practice, typical settings are advisable, which may also require some typographical knowledge.

Other confounding variables can emerge without considering higher order relationships, and these may not be recognised or their significance may be played down. For example, changing type size makes the line physically longer or shorter, if the number of characters is held constant. Also, the characteristics of standard visual display terminals available in the 1980s determined the choice and level of variables to some extent. However, even when screens permit finer control over typographic variables, there can be difficulties in interpreting the results because of confounding variables. These issues are explored in the analyses of individual studies in the following sections.

2. Line length

Line length can be measured by the physical length of the line, e.g. by adjusting margins. Some studies refer to this manipulation of physical length as visual angle, which takes into account the physical length and the viewing distance. Alternatively, line length can refer to the number of characters in a line. The number of characters per line can be varied by changing type size, but keeping the same physical length, e.g. 15 centimetres. This means that type size and number of characters per line are confounded. These relationships are illustrated in figure 1 in which each pairwise comparison results in one variable being held constant and two other variables changing. In comparing:

- (a) and (b), physical line length is constant, number of characters per line varies, type size varies;
- (a) and (c), number of characters per line is constant, physical line length varies, type size varies;
- (b) and (c), type size is constant, physical line length varies, number of characters per line varies.

In the studies described below, text is unjustified making the right margin uneven, so line lengths are generally based on average values.

2.1. Character density

In the 1980s, there were constraints on how type could be rendered on screen due to the display technology of the time. The two most common alphanumeric screens displayed 25 lines of 80 columns or 32 lines of 80 columns (Creed *et al.* 1987). A crude method of changing type size was to change the 'character density', which resulted in characters of the same height but different widths. This distortion of character shape introduces a further confounding variable, in addition to confounding type size and number of characters per line.

A study by Kolers *et al.* (1981) compared two character densities: a line length of 70 characters containing characters of half the width of a line length of 35 characters (illustrated in figure 2). Twenty participants read short extracts of about 300 words from Miller (1962) *Psychology, the Science of Mental Life*. Measuring eye movements, reading was found to be more efficient with the smaller characters, of which there were also more per line. With 70 characters per line there were more fixations per line, but the total number of fixations was fewer, the number of words acquired with each fixation was greater, the duration of each

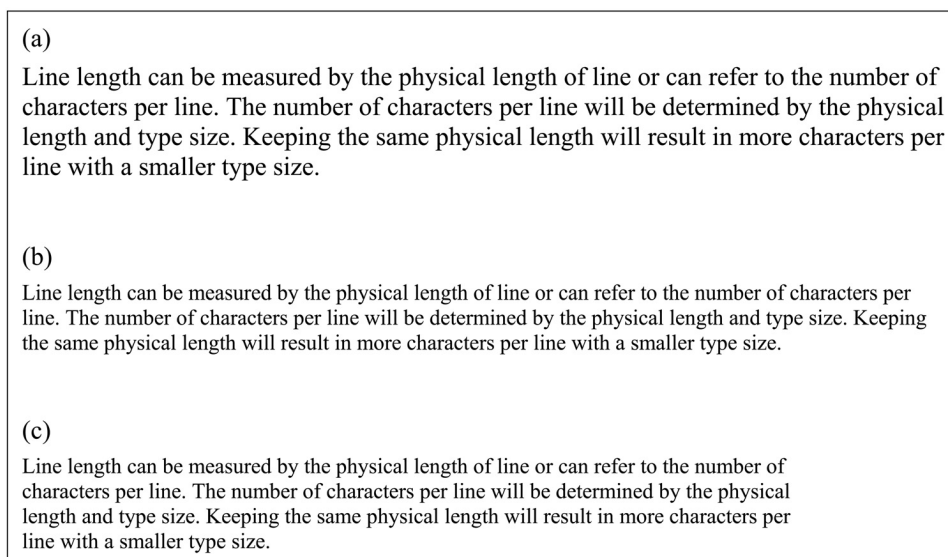


Figure 1. The relationship between physical line length, number of characters per line and type size: (a) and (b) both have a line length of 15 cm; (a) is set in Times New Roman 12 point resulting in approximately 90 characters per line; (b) is reduced to 10 point and therefore has approximately 100 characters per line; (a) and (c) both have approximately 90 characters per line, but their physical length varies because of the difference in type size.

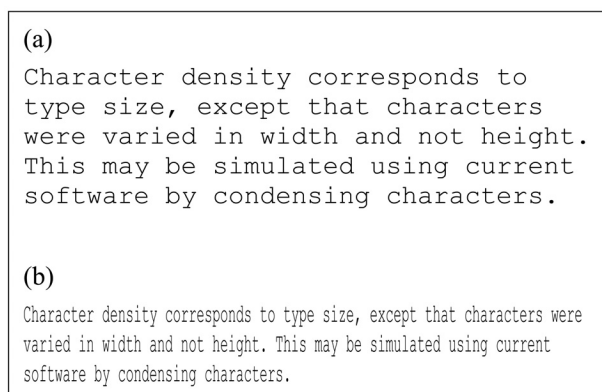


Figure 2. A simulation of different character densities used by Kolars *et al.* (1981). The monospaced characters in (a) have a line length of about 35 characters; the characters in (b) are half the width and result in a line length of about 70 characters (with the same physical length).

fixation was longer, and the overall reading time shorter. This experiment cannot identify whether smaller (narrower) characters or more characters per line produce more efficient reading.

2.2. Type size

Recent research, using a different generation of display technology screens with far more graphic

capabilities, has explored different type sizes (Bernard *et al.* 2002b). It can be implied from the report that the horizontal margins remained constant across conditions, thereby resulting in more characters per line with smaller type sizes, (a) and (b) in figure 1. The flexibility afforded by this technology would have enabled a comparison of type size, keeping the number of characters per line constant, which Gould and Grischowsky (1986) re-produced photographically (see later section on physical length). However, maintaining a constant number of characters per line confounds type size and physical length or margin size, (a) and (c) in figure 1. The question is then whether the number of characters per line or physical length is likely to have more of an effect on reading.

Bernard *et al.* (2002b) compared 10, 12, and 14 point sizes of eight type faces (Century Schoolbook, Courier New, Georgia, Times New Roman, Arial, Comic Sans MS, Tahoma and Verdana). Three groups of 20 participants read passages of slightly over 1000 words from Microsoft's Encarta, each group having a different type size. Participants were between 18 and 55 years old with an average age of 24 and the majority regularly read documents from screen. The passages contained substitution words and participants were instructed to read as quickly and accurately as possible and to identify the substitution words by stating them aloud. The substitutes were described as being 'inappropriate for the context of the passages' and varied grammatically. The example given in Bernard *et al.* (2001) is

substituting the adjective 'fake' for the noun 'cake'. Reading time and accuracy were combined (time/accuracy) and described as 'reading efficiency'.

The study found that fonts at the 10 point size were read significantly more slowly than fonts at the 12 point size, but there was a speed-accuracy trade-off removing differences between sizes. Fonts that were read faster were generally read less accurately. Participants' judgements of legibility were not determined by size alone but interacted with typeface. Increasing size did not necessarily improve the perception of legibility.

A larger character produces a larger retinal image, as does viewing the same size character from less distance. Morrison (1983) discusses the trade-off (within limits) between improved perceptibility of a larger retinal image and decreased perceptibility due to the image falling further into peripheral vision where acuity is poorer. This trade-off may explain the lack of effect of moderate differences in type sizes found by Bernard *et al.* (2002b) when accuracy of reading is taken into account. In print, 9, 10, 11, and 12 point type were found to be equally legible, measured by speed of reading with comprehension constant (Tinker 1965). However unlike Bernard *et al.* (2002b) the four type sizes were printed in different line lengths, which had been determined as optimal for the size of type (smaller type having a shorter physical line length than larger type).

2.3. Comparing character density and type size

The results of Bernard *et al.* (2002b) on type size differ from the comparison of characters of different densities tested using older screens (Kolars *et al.* 1981), where narrower characters with more characters per line were read faster. However, many differences between the two studies could account for the divergent findings. Those which are most closely related to the independent variable(s) are the technology and the range of stimuli. The older screen varied type size in a different way to current display technology. Also the number of characters per line in 14 point type is unlikely to be half the number of characters of 10 point type, so the Kolars *et al.* study is likely to have a larger range of characters per line. If number of characters per line is a variable which affects reading, there may need to be quite large differences before a change in reading speed is noticeable.

2.4. Physical length (visual angle)

Gould and Grischkowsky (1986) investigated the effect of visual angle on reading speed, co-varying

character size and physical line length. Eighteen experienced computer users participated in a proof-reading task, which was judged on reading speed and number of errors detected. The passages were excerpts from magazines of about 700 words. The authors' objective was to identify whether physically longer lines typically found on screen accounted for their slower reading compared to shorter lines in print. The experiment used photographs of screens and printed material, thereby assuming that other factors which are part of reading from screen do not interact with visual angle. As a consequence, these results should be regarded only as indicative of what might happen when reading from screen. They found that visual angle (or physical line length) has no effect on proofreading speed and accuracy over a middle range. The results also suggest no effect of character size within this range as the size of the characters within the experiment varied. However, at a small visual angle (where characters were very small) and a wide angle (large characters), reading rates were reduced.

2.5. Physical length and character density (number of characters per line)

A study carried out by Duchnicky and Kolars (1983) varied physical line length (measured in relation to screen width) alongside number of characters per line. By varying the amount of screen filled, through changing left and right margins, different visual angles are compared. A full screen width was compared with two-thirds and one-third screen width and the two character densities used by Kolars *et al.* (1981) were also included. Characters based on a 7 x 9 grid were displayed as light text on a dark background on a 30cm screen. Only ten participants were used, but each read 30 passages of about 300 words, again taken from Miller (1962) *Psychology, the Science of Mental Life*.

In this study, a check on comprehension involved saying whether questions could be answered by the text they had read. Reading rate was measured but not in the most obvious way, i.e. timing reading. Participants in the study adjusted their scrolling rate to read 'without either hurrying to catch up or waiting for a new line to appear' (Duchnicky and Kolars 1983: 686). Reading rate was based on the average scrolling rate (lines per second) of the last five complete lines to be read. This was converted to the number of characters per second, to account for the different number of characters in each condition.

The results confirmed the Kolars *et al.* (1981) study that smaller characters, with more characters per line, are read faster. The two-thirds and full screen widths

were also read faster than the one-third. Combining the two densities, a positive correlation indicated that the more characters per line, the faster the reading rate. There were no differences in comprehension.

The Duchnicky and Kolers experiment appears to vary physical line length independently of number of characters per line (as reported by Mills and Weldon 1987). However, partly due to the limitation of having only two very different character sizes, the effects of physical line length (width of screen filled) cannot be entirely separated from number of characters per line. Nor is it possible to separate the size of characters and number of characters per line, as was also the case with Kolers *et al.* (1981) and Gould and Grischkowsky (1986). The width of the wide characters constrain how many characters can fit on a line, so the three line lengths have relatively few characters per line. In contrast, the three line lengths of narrow characters span a much larger range of characters per line. However, an indication of the effect of characters per line, rather than character size or line length, can be made by comparing two data points. Narrow characters filling one-third of the screen and wide characters filling two-thirds of the screen both have 26 characters per line. No comparison is made within the article, but the average reading rates reported in their table 1 (22.4 and 20.5 characters per second) are similar, in the context of the range of reading rates across the experiment. This suggests that characters per line, irrespective of line length (screen width) or character width, may influence reading rate. Much more solid support for this proposal is the additional analysis reported in the article which found a positive correlation between characters per line and reading rate, referred to above. Both significant effects (character size and screen width) could be attributable to number of characters per line.

2.6. Characters per line

Using more recent display technology, Dyson and Kipping (1998a) looked at line length by keeping type size constant and varying the number of characters per line. This was done by changing the size of the right margin. Forty-eight participants, the majority between 25 and 34 years old and fairly frequent computer users, read passages of about 800 words taken from the Microsoft Network. Reading rate and comprehension were measured and six line lengths compared, ranging from 25 to 100 characters per line. The text was set in 10 point Arial with a consistent interlinear spacing of 12 point. Three of the six line lengths are shown in figure 3.

The results showed that the number of characters per line affects reading rate with the longest line read faster

than the shortest line tested. Although there was a visible trend towards faster reading as the number of characters increased, there were no other statistically significant differences.

By calculating the amount of time necessary to scroll through documents with different numbers of characters per line (and hence number of lines), Dyson and Kipping were able to estimate if scrolling time completely accounted for differences in reading rates. Faster reading times for the longest line could be partly attributed to the reduction in scrolling time required to move through the document. However, this factor did not seem to completely account for the differences found. They concluded that participants sometimes read whilst scrolling and reading patterns appear to be adjusted according to the number of characters per line and the conditions of the experiment.

Another possible explanation for faster reading at 100 characters per line is that this line length is less affected by glare from the screen. Line lengths that did not fill the screen had a larger right margin which was an area of blank (white) screen to the right of the text, which may have proved distracting. A follow-up experiment was therefore conducted with 25, 55 and 100 characters per line placed within text windows with a background of grey to the right of the windows. This was designed to avoid any interference with reading caused by glare. Reading was again faster with 100 characters per line than 25, but this difference could have been attributable to scrolling time. Reducing glare may have helped the reading of shorter lines.

Bernard *et al.* (2002a) also compared line lengths using 45, 76 and 132 characters per line by varying left and right margins. Twenty adults, with an average age of 29, and who read from screen a few times a week read passages of about 1000 words taken from Encarta. These were displayed in 12 point Arial, but the interlinear spacing is not specified. The research found no differences in reading times or reading efficiency (time/accuracy) using the method of reading and identifying substitution words described earlier. Their interpretation of these results is that the positive effects of fewer characters per line (shorter line lengths) were offset by the longer scrolling time, the assumption being that shorter line lengths are better.

The authors of this study offer the explanation that scrolling time contributes to differences in reading rates from screen, but they provide no evidence for this. The study also fails to replicate Dyson and Kipping (1998a) and earlier studies that more characters per line can result in faster reading. The difference may be due to the extreme nature of the longest line, i.e. 132 characters in 12 point Arial (rather than 10 point Arial used by Dyson and Kipping). The line length therefore not only has

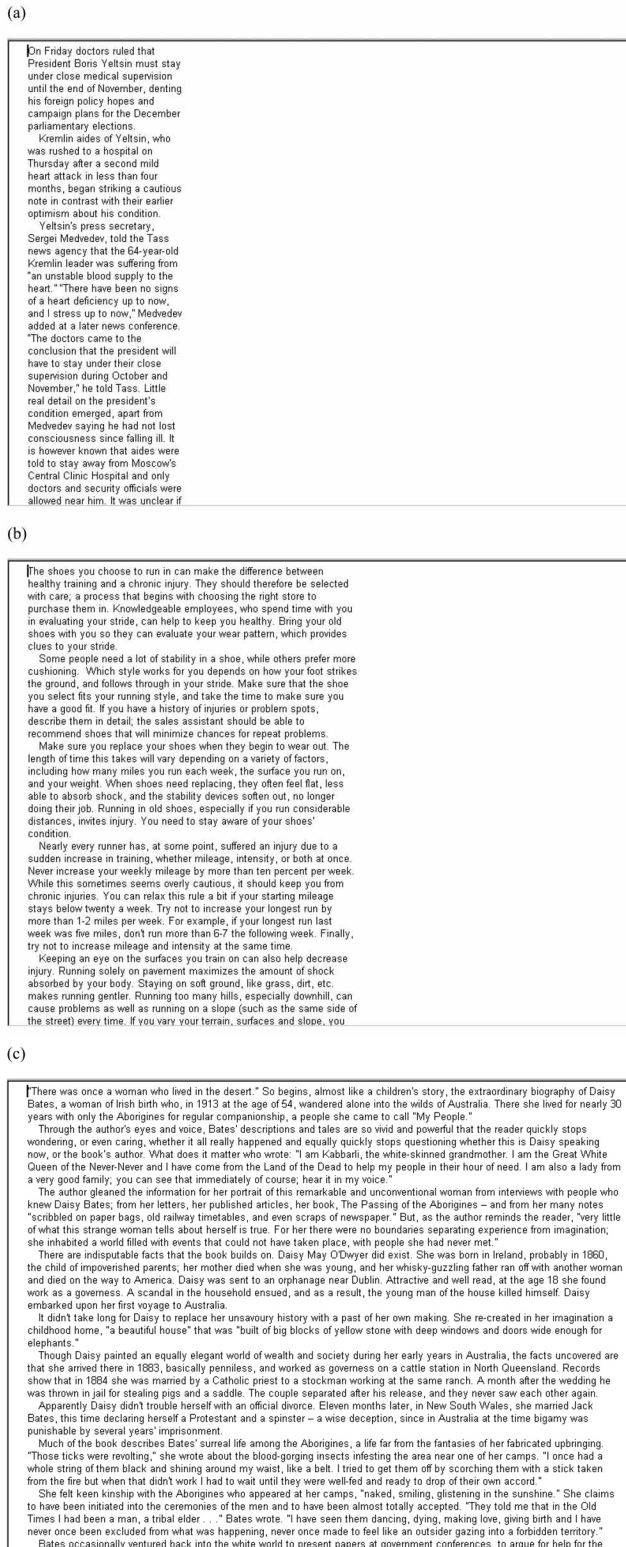


Figure 3. Line lengths of approximately (a) 25, (b) 55 and (c) 100 characters per line with variation in the size of the right margin.

more characters but is also physically longer because of the larger type size. Another difference between the two recent studies is the positioning of the text. In Dyson and Kipping (1998a) this was on the left of the screen, varying the right margin, whereas Bernard *et al.* (2002a) positioned text in the centre of the screen, varying the size of both left and right margins.

2.7. Characters per line and margins

A study which includes margins as an independent variable, not confounded with line length, was carried out by Youngman and Scharff (1998). They compared three physical line lengths, (4, 6, and 8 inches), and four sizes of margin (0, 0.5, 1, or 1.5 inches). A setting with no margin would not be typical practice, but may have been included to assess an extreme of a variable in a similar manner to using 132 characters per line. As type size remained constant, the number of characters per line must have varied, but these are not reported. The typeface used was 12 point Times New Roman in black on a grey background, chosen because it was said to be the default for several Internet browsers.

Twenty-seven participants scanned excerpts of text for hidden words, which could be star, square, diamond, circle, or triangle. They responded by clicking on the word shape at the bottom of the screen. The measure used was the median reaction time for correct responses. There was a significant interaction between text width (characters per line) and margin size, based on the three text widths and 0 and 0.5 margins. Of particular interest is the fact that the fastest reaction time was with the longest line length and hence largest number of characters per line (and no margin). Based on information provided in Youngman and Scharff's report, this line length is estimated to be about 100 characters and is unlikely to be as long as 132 characters.

2.8. Comprehension of characters per line

Few studies specifically address the effect of line length on comprehension. It is more common to aim to keep comprehension constant and look for differences in reading rate. For example, Dyson and Kipping (1998a) used Duchnicky and Kolers' comprehension test which is a check on comprehension, rather than a task which aims to detect differences between conditions, and found no differences in comprehension. However, Dyson and Haselgrove (2001) found that line length influences readers' comprehension with documents at 55 characters per line producing better comprehension scores than the longest line length (100 characters per line). A more

elaborate test of comprehension was used in this study. A set of multiple choice questions requiring recall of different types of material was developed by Dyson and Haselgrove (2000) and used in the 2001 study to compare line lengths. Differences in comprehension were not cancelled out by differences in reading rate as there was no speed-accuracy trade-off.

2.9. Subjective judgements of characters per line

Readers' perceptions of different line lengths have also been measured. Grabinger and Amedeo (1988) asked participants to judge the 'study-ability' of computer-generated text types, meaning the ease with which text could be read and studied. The research aimed to inform the design of text on screen, but used printed examples in this study. The experiment was misleadingly described by de Bruijn *et al.* (1992) as judgements of the layout of text presented on screen. Amongst other variables (such as interlinear spacing, paragraph denotation, heading location) they compared two line lengths of 40 and 60 characters per line. They found that shorter lines contributed to judgements of better organization and simplicity.

This approach of using paper models to derive empirical evidence for the effects of typographical variables on screen should be questioned. However, this criticism was subsequently addressed by Grabinger (1993) by using screens and further explored by comparing 'model' and actual screens. The actual screens were copied from existing programs and varied in content and design characteristics, ensuring the study was ecologically valid. Model screens controlled for content effects by adapting the notation developed by Twyman (1981), using x and o to represent text. Similar results were found for the two types of screens and these also supported and refined the earlier results from paper.

Dyson and Kipping (1998a) measured readers' perceptions of ease of reading by asking participants to compare screens. Their perceptions did not correlate with their performance. A medium line length of 55 characters per line was rated as easiest to read, but was not read the fastest. Similarly, Youngman and Scharff (1998) found that the longest line length was disliked the most, despite good performance.

A more unusual question was asked of participants by Bernard *et al.* (2002a). They were asked to judge whether the amount of scrolling was optimal for the particular line length. The greatest number of characters per line was considered as most optimal in terms of the amount of scrolling. Put more simply, this seems to suggest a preference for less scrolling.

3. Columns

Presenting text in columns provides a different way of comparing line lengths. Efficient use can be made of space by combining shorter line lengths and a multi-column format. Newspapers and magazines which typically use such formats are available on the WWW. In terms of empirical studies, introducing multiple columns allows short line lengths to be compared with longer line lengths with a similar amount of text per screen. However, few studies have compared single and multiple columns on screen and the tasks differ, making it difficult to draw conclusions.

3.1. One vs. two columns

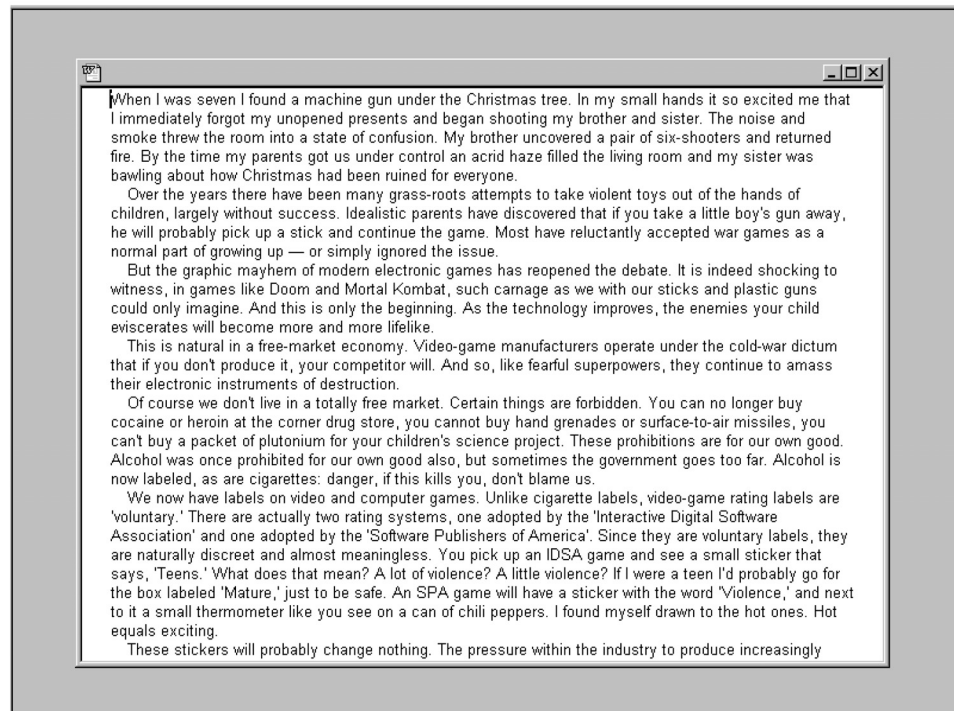
Creed *et al.* (1987) included a two column format in their comparison of proof reading on screen and paper, to see whether performance could be facilitated on the screen. They used texts which filled three screens and compared two columns of 30 characters wide and a separation of 10 characters with one column of 70 characters; both contained 25 lines. There was no effect of column format, but a suggestion that there may have been a speed-accuracy trade-off with two columns. For screen and paper, reading was slightly faster in two columns, but fewer errors were detected.

3.2. One vs. three columns

A comparison of one and three columns was made by Dyson and Kipping (1997) measuring reading rate and comprehension. The texts were chosen to be typical of what might be found in online magazines, with passages of a little under 2000 words taken from *New Scientist*. These were displayed in 10 point Arial with 12 point interlinear spacing. Eighteen participants each read texts in three conditions: a single column (about 80 characters per line) with either scrolled or paged movement and a three column paged format (about 25 characters per column). Figure 4 illustrates the formats. Paging is a more suitable option for three columns as columns can only practically be read to the bottom of the screen. The single paged column was read faster than the other two formats, with no differences in comprehension. This confirms the advantage of longer lines reported earlier, and also provides a direct measure of the effects of scrolling on reading rate.

However, further exploration of this data showed that faster reading of the single paged column was only found with the younger age group tested (18–24 year olds). Those over 25 years old showed no differences in

(a)



(b)

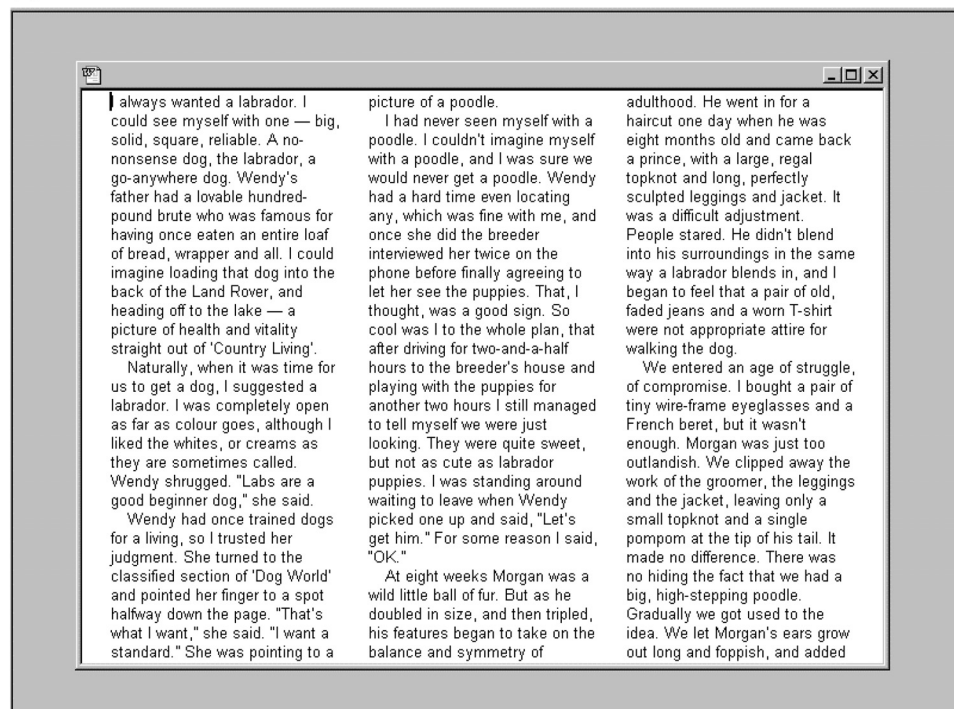


Figure 4. (a) A single column of approximately 80 characters per line (presented with both scrolling and paged movement); (b) three columns of approximately 25 characters per line (with only paged movement).

reading rate across the three formats. This result is difficult to explain other than proposing that partici-

pants' familiarity with the formats may contribute to speed of reading. The single column was designed to

simulate a typical web page, which may have been more familiar to the younger group. Although all participants (except one) were relatively experienced computer users, the type of activities in which they engage may differ. The paged format may have been read faster because participants did not need to scroll through the text. However, this would apply to both age groups and therefore does not fully explain the results.

The relatively limited amount of studies into single and multiple column formats in print cannot help in interpreting these findings for screen. They have produced inconclusive results, but this is unsurprising as many variables (independent and dependent) other than the line length differ among experiments. For example, an advantage for narrow column setting was found by Foster (1970); Poulton (1959) reported that single columns were read faster than double columns; no differences were found between a single and two column setting by Hartley *et al.* (1974).

The comprehension task used by Dyson and Kipping (1997) was a check on comprehension used by Kolers *et al.*, and discussed earlier. Generally, no differences in comprehension have been found using this method (Dyson and Kipping 1998a,b). However, a difference in comprehension across the three formats was found when readers were divided into faster and slower readers (Dyson and Kipping 1997). In the three column paged format, comprehension was better for faster readers than slower readers. Reports of better comprehension for faster readers have come from Tinker (summarised in Tinker 1963) and Jackson and McClelland (1979). However, the reason for this differentiation with the three column format needs to be considered. Dyson and Kipping (1997) propose that different types of readers are adopting different reading patterns that affect their comprehension. A faster reader may be able to scan narrow columns in an efficient manner to absorb material.

This explanation has not been tested directly, but Dyson and Haselgrove (2001) did compare comprehension of different line lengths at normal and fast reading speeds. (Participants were trained to read fast, rather than separating readers into faster and slower readers). There was no interaction between reading speed and line length, but as reported earlier, better comprehension with shorter (55 characters per line) rather than long (100 characters per line) lines. The shortest line (25 characters per line) equivalent to the three column width, fell in between. Therefore, there does not seem to be an advantage for short lines when reading fast. However, as the short line length was not set in three columns, and naturally fast readers were not identified, the strategy proposed for fast readers of narrow columns has still to be tested.

3.3. Subjective judgements

Subjective ratings are in favour of more than one column. Dyson and Kipping (1997) found that the three column format was perceived as easiest to read. Asking for judgements of how much they would like to read or study from screen, Grabinger (1993) found that two text columns contributed to more organised and more visually interesting texts.

4. Window size

Empirical studies have also looked at the vertical dimension of screens, varying window height by changing the number of lines per screen. This variable is not particularly relevant to the legibility of print, as there is generally less constraint on the length of a page.

4.1. Few lines vs. 20 lines per screen

Five window heights of 1, 2, 3, 4 and 20 lines were compared by Duchnicky and Kolers (1983) in the study described earlier in relation to line length. Reading was faster with 4 and 20 lines compared to fewer lines, but these two very different sizes were not different in terms of reading rate. There were no differences in comprehension across all five window heights.

4.2. 20 vs. 40 lines per screen

Kruk and Muter (1984) also manipulated the number of lines, but only in their 'book' condition, not on screen. Overall they compared three conditions: two in print (20 and 40 lines per page) and one on screen (20 lines per screen). The characters were low resolution, based on a 5×7 grid, and were displayed as green text on a black background on a 30.5 cm monitor. The printed version used a Pica 10 pitch daisy wheel. Twenty-four undergraduates read short stories 'as if for pleasure' for periods of 5 min and the extent of reading was measured.

The study indicated that 40 lines were read faster than 20 lines in either print or on screen and there were no differences in comprehension scores. However, as the difference in reading rate between the two print conditions was relatively small compared with the difference between print and screen, they concluded that the format could not completely account for differences between reading from paper and screen.

There are two problems with this interpretation of the results. The aim of the study was to explore reasons for

slower reading of text from screen, rather than the effect of different screen formats. However, this experimental design assumes that changing the number of lines in print will have the same effect on screen. This is not a suitable method to collect empirical evidence for the effects of typographical variables on screen. Also, the difference in reading speed might be due to the number of characters per line as these differed between conditions. With 40 lines per page there were more characters per line (60 as compared to 39).

4.2. 20 vs. 60 lines per screen

Looking at comprehension and reading rate on screen, Dillon *et al.* (1990) used a longer text (3500 words) than Duchnicky and Kolers (300 words) and compared screens of 20 and 60 lines. This study used a high resolution A3 screen with black text on a white background, but no further description of typographic variables is given. Thirty-two participants with a mean age of 29 read the texts for understanding, with no time pressure.

No differences in performance (reading rate or comprehension) were found between the two sizes. However, they did find that participants using the small screen jumped about more between 'pages' and changed direction of viewing more than those reading from the larger screen. These manipulations took account of the need to display more pages with a smaller screen. Dillon *et al.* (1990) point out that the suggestion from the Duchnicky and Kolers (1983) study that around five lines is optimal is unlikely.

4.4. 23 vs. 60 lines per screen

A different outcome from reading screens of two sizes is reported by de Bruijn *et al.* (1992). The larger screen measured 15 inches (approximately 38 cm) and displayed 60 lines of text with characters based on a 9×11 grid. The smaller screen was a 12 inch (30.5 cm) display of 23 lines with characters based on a 9×16 grid. A second variable was the structure of the layout which is described in a later section of this paper (typographic layout). Fifty-six students spread over four conditions read a 'legal-sociological discourse' of approximately 1900 words and were instructed to understand and remember the text.

They found that 'learning time' was greater for a screen containing 23 lines than one with 60 lines. Other performance measures (summary and multiple choice test; reaction time to secondary task) were not affected by window size. The authors offer some possible

explanations for the divergence between their results and those of Dillon *et al.* (1990). These include physical aspects of Dillon *et al.*'s screen and the possibility that different strategies were adopted by participants in the two experiments. However, a specific problem with the de Bruijn *et al.* (1992) study is the number of variables which differed between the two conditions because two different models of screens were used. These variables are referred to in their discussion as refresh rate, resolution and character size, but the authors consider their effects to be minimal.

Another difference between the two screens which may have contributed to the learning time difference is the time spent in physical manipulations. Unlike Dillon *et al.* (1990), they did not take account of the need for more downward movements resulting from a small screen. It is possible that the time involved in these movements may have contributed to the longer learning times, due to time spent in selecting the action and the repeated refreshing of the screen.

4.5. 15, 25, 35 lines per screen

Using differences in window sizes, and a screen technology that removes differences in refresh rates depending on display size, Dyson and Kipping (1998b) compared 15, 25 and 35 lines per screen. Twenty-four participants read documents of about 700 words in Arial 10 point with 12 point interlinear spacing and a line length of 55 characters per line. No differences in reading rates or comprehension were found.

4.6. Other tasks

The influence of the number of lines of text on some other tasks has also been investigated. Richardson *et al.* (1989) introduced an information location task with text presented in window heights of 20 and 40 lines. Sixteen participants were used but no other details concerning the material are provided in this paper. As with the measures of reading rate and comprehension, they found no difference between the two sizes in time to complete the task.

The ergonomics of reading comprehension were explored by Askwall (1985) using a reasoning task which required judgements as to the validity of inferences. Seventeen students with little or no previous experience with reading from screen read fictional texts of about 150–200 words. Information was presented either on one screen (displaying 24 lines of 40 columns) or sentence-by-sentence across a number of screens. This manipulation varies the number of lines per

screen, although these will not be consistent from screen to screen. Participants read more slowly in the sentence-by-sentence condition. Breaking down the total time into time spent searching the text and time needed for reading and processing the text showed that slower searching within sentences accounted for the difference. This may be explained by the mechanics of the interface, i.e. additional time would probably be required to access more individual screens. Therefore, as with the above studies, this result does not seem to provide support for an effect of number of lines on reading rate.

Proof reading has also been explored with different text formats. Creed *et al.* (1988) initially compared a single paragraph (between 4 and 15 lines) with an entire screenful (25 lines) and then included a sentence (between 1 and 8 lines). In each of two experiments, 24 undergraduate participants read passages of around 800 words (filling just over three full screens) and text was displayed as 25 lines of 70 characters. Although the results were not entirely consistent across the two experiments reported within the paper, there seems to be a tendency for slower reading of smaller amounts with correspondingly greater accuracy. This speed-accuracy trade-off essentially removes the effect of the number of lines on performance. However, as Creed *et al.* (1988) point out, since proof reading generally aims to achieve maximum detection rate for errors, these results recommend using sentences or paragraphs, rather than whole screens.

4.7. Subjective judgements

Although window size may not affect reading rate and comprehension, subjective judgements of window size appear to match our expectations more closely, as with line length. In both the Richardson *et al.* (1989) and Dillon *et al.* (1990) studies, participants expressed a preference for the larger window. Similarly, the smallest window size (15 lines) presented by Dyson and Kipping (1998b) was judged as least easy to read. De Bruijn *et al.* (1992) used a more detailed set of questions to obtain participants' opinions of the quality of text layout and 'cognitive-ergonomic' aspects of screens. They found no effect of screen size for either dimension, but this may be partly attributable to the inclusion in the experiment of another independent variable. A well- and ill-structured text layout was included at both screen sizes. This factor affected perceptions of clarity, organisation, ease of reading and comfort. Where more than one factor is manipulated in a layout, subjective ratings may be dominated by one of these. Line length

can influence judgements of ease of reading more than scrolling vs. paging (Dyson and Kipping 1998a). The amount of text per screen (sentence vs. paragraph) can affect preference more than contrast polarity (Creed *et al.* 1988).

5. Interlinear spacing

Interlinear spacing can also be used to vary the amount of text on screen, but is usually considered in relation to how much space there is between lines. This variable is often overlooked (i.e. not specified) in reports of studies varying other typographic features (e.g. Bernard 2002a).

5.1. Single vs. double spacing

Kolers *et al.* (1981) looked at interline spacing, comparing single and double spacing in addition to character density. They found that single spacing required a few more fixations per line, slightly fewer words were read per fixation, and total reading time was slightly longer. They conclude that double spacing is marginally superior to single spacing.

5.2. Single vs. double spacing and lines per screen

Kruk and Muter (1984) separated the effects of the number of lines per screen from the interlinear spacing by using only part of the screen for single spacing and keeping number of lines constant. They found that reading was significantly slower with single spacing compared to double spacing, a larger effect than that reported by Kolers *et al.* (1981). According to Kruk and Muter, this discrepancy may be due to isolating spacing as a variable, removing the confounding with lines per screen. Identifying this confound is useful however, subsequent studies of the effect of lines per screen (described above) cast doubt on this explanation. If there is any effect, it is that fewer lines per page (double spacing) are read slower. An alternative possibility is that the single spacing condition may have had slightly more space between the lines in the Kolers *et al.* (1981) experiment.

Grabinger (1993) included interlinear spacing as one of eight variables used to construct screens for comparison. This study is described in the next section.

More recent display technology permits a finer grained measurement of interlinear spacing than single and double spacing, but there do not seem to have been recent empirical studies of this variable.

6. Typographic layout

Some studies have taken a global approach to layout by comparing layouts which vary a number of features. As discussed earlier, this approach acknowledges that the relationship between typographic variables is important. When developing teaching material, Grabinger (1993: 36) describes the role of a screen designer as 'devising a combination of format variables to enhance reading and studying'.

Two main approaches are represented in the studies below: one which seeks to identify the dimensions underlying judgements of layouts and another which contrasts 'good' and 'poor' layouts and measures their effects. The latter approach has been implemented with printed material where re-designs (involving many variables) have been shown to improve effectiveness (Hartley 1994).

6.1. Identifying dimensions for making judgements

Grabinger (1993) used the first approach, combining variables in order to identify constructs to inform the design of screens. As described earlier, this study used model screens (with no content) and these varied their layout according to eight variables. Of particular relevance to this paper are interlinear spacing (single or double) and columns (single or double). The combination of eight variables, each with two levels, produces 256 possible screens, an impractical number to test. Therefore not all combinations were used and each of the 94 participants, ranging in age from 20 to over 60, made 62 paired comparisons judging the readability and studyability of the screens.

The results indicated that the organisation of screens and their visual interest were relevant dimensions and single spacing and double columns were among the characteristics associated with more positive judgements. However, as other characteristics were also varied, the individual contributions of interlinear spacing and columns cannot be identified.

6.2. Well-designed and poorly-designed

A relatively early study (Duin 1988) comparing two versions of screen designs was carried out within the context of computer-assisted instruction. The design of the materials was based on conceptual, linguistic, visual and movement guidelines. The visual guidelines are relevant to this paper and these included keeping text density low, right margins unjustified, and putting blank lines between paragraphs. Twenty-nine college

students worked with the well-designed material on screen and 29 with the poorly-designed material which produced cluttered screens. As the study was concerned with instructional material, the measures taken aimed to reflect the learning process, rather than efficiency of reading. These looked at how students worked with the material and the quality of their writing, but also included students' subjective judgements.

In comparison to the poorly-designed version, the well-designed version appeared to help students as they asked fewer questions, worked through the task with greater ease, were less anxious and showed greater flexibility. These observations were made by trained independent observers recording behaviour. This version also resulted in higher quality writing.

As this study manipulated the content and nature of the material, as well as its layout, it is not possible to isolate the effects of layout. However, a question which specifically asked students about the visual layout found that they judged the well-designed material more favourably.

6.3. Enhanced and standard format

Muter and Maurutto (1991) compared two screen formats: an 'enhanced' and a 'standard' format, with a 'book' format. Nine enhancements are listed, two of which have been addressed in this paper (line length and interlinear spacing). Lines were a maximum of 80 characters and double spaced. Every other line was indented by three spaces based on a recommendation by Huey (1908) that this facilitates return sweeps. The standard format was designed to resemble personal computer screens of the 1980s and differed on the variables mentioned by being single spaced and having no indentation of every other line. There was no difference in reading rate or comprehension between the conditions (including the 'book'). However, 13 of the 18 participants preferred the enhanced format compared to the standard.

As Muter and Maurutto (1991) discuss, the lack of an effect on performance of text format may have been due to some of the supposed enhancements having a negative effect. One in particular seems to be a strong candidate. Indenting every other line is an unusual way of presenting text, unfamiliar to most readers. Surprisingly, this does not appear to have influenced participants' preference judgements. However, their performance may have suffered as more recently than Huey (1908), investigations into the visual process of reading have suggested that the left-hand margin should be justified (Bouma 1980).

6.4. *Well-structured and ill-structured*

A further study which included two text layouts was carried out by de Bruijn *et al.* (1992) and referred to when discussing subjective judgements of window height. These authors used the same method as Muter and Maurutto for devising a 'well-structured' text in that previous research or guidelines informed design decisions. However, de Bruijn *et al.* (1992) appear to go further than modifications to the presentation by including, for example, 'first', 'second' in the left margin of the well-structured text to order reasons and arguments. The results are similar to Muter and Maurutto (1991) in that they found no differences in performance, but differences in subjective judgements. Participants judged the well-structured text layout as clearer, better organised and easier to read than the ill-structured format.

7. Discussion

Most of the studies on line length report faster reading with longer lines, and point to the number of characters as the variable responsible for the differences, rather than physical line length (visual angle). Increasing characters per line, but maintaining a constant visual angle can result in faster reading (Kolers *et al.* 1981, Duchnick and Kolers 1983), whereas differences in visual angle have only a small effect on reading speed within the middle range applicable to most displays (Kruk and Muter 1984, Gould and Grischkowsky 1986).

When type size is constant within a study, the size of margin will be confounded with number of characters per line (e.g. Dyson and Kipping 1998a, Youngman and Scharff 1998, Bernard *et al.* 2002a). The Youngman and Scharff (1998) study suggests that size of margin may be relevant, but the exact nature of their stimulus material is unclear. Dyson and Kipping (1998a) suggest that margins may have some effect through the colour of the blank background. The precise contribution of the location, size and background colour of margins to reading performance has not yet been determined, which suggests that thought should be given to how this variable is treated in future studies. However, the converging evidence indicates that the number of characters per line is a more important variable.

The number of characters is the appropriate metric for saccade size. An empirical demonstration carried out by Morrison and Rayner (1981) confirmed that saccade size is consistent in terms of number of characters, and not visual angle. Eye movements measurements by Kolers *et al.* (1981) found there were more fixations per line with more characters per line, in keeping with a

consistent number of characters per saccade (Morrison and Rayner 1981).

However, increasing characters per line and maintaining a constant visual angle (Bernard *et al.* 2002b) and increasing characters per line along with visual angle (Bernard *et al.* 2002a) may not result in faster reading. This may be because there is a range of line lengths within which changes to the number of characters have little effect. Going outside this range, with short or long lines, may affect reading speed, with slower reading of short lines and faster reading of lines as long as 100 characters (Dyson and Kipping 1998a).

This pattern of results differs from those found with print, where Spencer (1968) states that line lengths should not exceed about 70 characters per line. Rayner and Pollatsek (1989) have interpreted the results from studies by Tinker and co-workers (summarised in Tinker 1963) and come to a more specific conclusion on line length. They deduce that Tinker's work identified an optimal line length of 52 characters per line.

Both return sweeps to the beginning of the next line and fixations within a line can be affected by line length. The explanation given for the legibility of a moderate line length in print is that it is the outcome of a trade-off between two opposing factors. If the lines are too short, readers cannot make use of much information in each fixation. If line lengths are too long, the return sweeps to the beginning of the next line are difficult.

There appears to be no evidence to suggest that eye movements when reading from screen are different from reading print (Gould *et al.* 1987). However, we tend to sit further away from the screen than from printed matter when reading (Gould *et al.* 1987), so that a longer physical line length on screen subtends a smaller visual angle compared to the same physical line length in print. Reading of long lines on screen does not appear to be slowed down, so it is possible that reading from screen presents less of a problem with return sweeps to the next line, as these cover a smaller visual angle. Without this difficulty, there may be a saving in time spent in eye movements of long lines as the total time in return sweeps is theoretically decreased (Dyson and Kipping 1998b). The relationship between the amplitude (A) of saccades (e.g. return sweeps) and their duration can be described by the function $2.2A + 21$ ms (Carpenter 1977), which means that the greater number of return sweeps with shorter lines will add more to the time than longer lines.

As the difference in time in return sweeps between 25 and 100 characters per line does not provide a complete account of the observed differences (Dyson and Kipping 1998b), other factors specific to reading from screen should also be taken into account. One of these is the mechanics of scrolling which can contribute to faster

reading of longer lines on screen. With more characters per line, less time is necessarily spent in scrolling as the document is shorter (in terms of numbers of lines). However, this explanation does not appear to completely account for differences.

Scrolling through the text may also provide a cue to locating the next line. Duchnicky and Kolers (1983) suggest that it may be the upward movement of scrolled text which reduces the difficulties associated with longer lines. This explanation has not been explored empirically.

There is likely to be a point beyond which increasing the number of characters per line no longer improves reading rate, which may have been demonstrated by Bernard *et al.* (2002a) using 132 characters per line. Such a large visual angle may have caused significant disruption to the return sweep, perhaps even requiring head movements as well as saccades. Studies to date have not identified a limit and whether it depends on visual angle or number of characters. In increasing line length with more characters per line, the advantages of fewer return sweeps and more fixations per line may be offset at some point by difficulties in locating the next line. Increasing interlinear space might help in more accurate return sweeps, but this has not been explored and limited studies from the early 1980s are inconclusive.

These explanations for faster reading of longer lines on screen do not address comprehension. In most studies, comprehension is included only to check whether there is a speed-accuracy trade-off (e.g. Bernard *et al.* 2002b). However, there are some suggestions that line length may affect comprehension and this may be related to the mechanics of scrolling and reading patterns (Dyson and Haselgrove 2001).

If the effects of line length are related to the mechanics of reading (eye movements and scrolling), changing the amount of text on screen through the number of lines would be unlikely to influence reading. Although there are differences in how the text is manipulated, and larger windows are perceived as better, there is no reliable evidence that reading rate or comprehension are affected.

Studies which considered layout and other factors as a whole have, to some extent, validated their choice of settings through the consistency of subjective judgements in favouring the 'good' designs. However, it is difficult to find evidence for a layout which supports more efficient reading from screen. Either other factors may be responsible or there are no performance differences. Whilst significant differences in performance would still have led to questions as to which typographic variable(s) and/or interaction between variables might be responsible for the differences, no differences

unfortunately offers no direction for further investigation.

In general, subjective judgements of variables relating to text format on screen have not been in close agreement with objective performance measures, such as reading rate and comprehension. Although longer line lengths may be read faster, people prefer a more moderate length. With columns, a single wide column is read faster, but narrow multiple columns are preferred. With window height, the only performance differences relate to manipulating text, but larger windows are perceived as better. Performance measures in relation to type size are different from perceived legibility.

This general trend can also be found with research into the legibility of print. Spencer (1968) concludes that there is little correlation between preferences or opinions of readers and objective measures of legibility. Some subjective judgements of text on screens may be influenced by what is typically experienced in reading certain types of printed material (e.g. 50–70 characters per line). This experience will change over time which may result in formats that are frequently used on the web being perceived in a positive way. Whether this familiarity also leads to improved performance requires investigation.

8. Concluding remarks

With current interfaces, reading from screen can be investigated by manipulating text format in numerous ways. It is inevitable that some of these manipulations will introduce confounds between typographic variables, but some confounds may be more damaging to the validity of the results than others.

This synthesis of studies comes to the conclusion that the number of characters per line is an important variable which can affect speed of reading. Therefore, in investigating for example, type size, line length in characters per line should either remain constant or be systematically varied. Attempting to use optimal configurations (see Tinker 1965) or 'good' layouts (Muter and Maurutto 1991, de Bruijn *et al.* 1992) not only relies on interpreting previous research combined with skill in designing text material, but also has limited theoretical significance.

Empirical research on reading from screen has spanned more than 20 years, but progress in developing a sound body of knowledge on the effects of text formats is slow. This is likely to be due in part to changes in technology, requiring studies which attempt to replicate, rather than extend results. In addition, there may be a relative lack of interest in such research, if the outcomes are considered predictable from research into print.

In this paper, comparisons with the findings from reading print suggest that this assumption is misleading, and issues for further investigation can be identified. These include: finding the line length at which an increase in characters per line slows down reading and the factors responsible; exploring how scrolling movements contribute to reading efficiency; looking at how individual reading styles or patterns interact with text format; evaluating whether the familiarity of text format affects perceptions and performance (through practice as well as experience).

Other areas that were defined as falling outside the scope of this paper have a research literature for print, but this is limited in relation to screen. Issues relating to the structure of text on screen, the use of space and other devices, have yet to be systematically investigated. In addition to the reasons already proposed for the limited number of studies, a more positive explanation may be the diversion of resources into researching navigation on screen. Electronic documents enable multiple navigation routes and different ways of reading, which produce a set of research questions for reading from screen that do not have a parallel in reading print.

References

- ASKWALL, S. 1985, Computer supported reading vs. reading text on paper: a comparison of two reading situations. *International Journal of Man-Machine Studies*, **22**, 425–439.
- BERNARD, M., FERNANDEZ, M. and HULL, S., 2002a, The effects of line length on children and adults' online reading performance. *Usability News*, **4**, http://psychology.wichita.edu/surl/usabilitynews/42/text_length.htm
- BERNARD, M., LIDA, B., RILEY, S., HACKLER, T. and JANZEN, K. 2002b, A comparison of popular online fonts: which size and type is best? *Usability News*, **4**, <http://psychology.wichita.edu/surl/usabilitynews/41/onlinefont.htm>
- BERNARD, M., MILLS, M., PETERSON, M. and STORRER, K. 2001, A comparison of popular online fonts: which is best and when? *Usability News*, **3**, http://psychology.wichita.edu/surl/usabilitynews/3S/usability_news.html
- BORCHERS, J., DEUSSEN, O., KLINGERT, A. and KNÖRZER, C. 1996, Layout rules for graphical web documents. *Computers & Graphics*, **20**, 415–426.
- BOUMA, H. 1980, Visual reading processes and the quality of text displays. *IPO Annual Progress Report*, **15**, 83–90.
- CARPENTER, R.H.S. 1977, *Movements of the Eye* (London: Pion).
- CREED, A., DENNIS, I. and NEWSTEAD, S. 1987, Proof-reading on VDUs. *Behaviour & Information Technology*, **6**, 3–13.
- CREED, A., DENNIS, I. and NEWSTEAD, S. 1988, Effects of display format on proof-reading with VDUs. *Behaviour & Information Technology*, **7**, 467–478.
- DE BRUIJN, D., DE MUL, S. and VAN OOSTENDORP, H. 1992, The influence of screen size and text layout on the study of text. *Behaviour & Information Technology*, **11**, 71–78.
- DILLON, A. 1992, Reading from paper versus screens: a critical review of the empirical literature. *Ergonomics*, **35**, 1297–1326.
- DILLON, A., RICHARDSON, J. and MCKNIGHT, C. 1990, The effects of display size and text splitting on reading lengthy text from screen. *Behaviour & Information Technology*, **9**, 215–227.
- DUCHNICKY, R.L. and KOLERS, P.A. 1983, Readability of text scrolled on visual display terminals as a function of window size. *Human Factors*, **25**, 683–692.
- DUIN, A.H. 1988, Computer-assisted instructional displays: effects on students' computing behaviors, prewriting, and attitudes. *Journal of Computer-Based Instruction*, **15**, 48–56.
- DYSON, M.C. 1999, Typography through the eyes of a psychologist. *Hyphen*, **1**, 5–13.
- DYSON, M.C. and HASELGROVE, M. 2000, The effects of reading speed and reading patterns on our understanding of text read from screen. *Journal of Research in Reading*, **23**, 210–223.
- DYSON, M.C. and HASELGROVE, M. 2001, The influence of reading speed and line length on the effectiveness of reading from screen. *International Journal of Human-Computer Studies*, **54**, 585–612.
- DYSON, M.C. and KIPPING, G.J. 1997, The legibility of screen formats: are three columns better than one? *Computers & Graphics*, **21**, 703–712.
- DYSON, M.C. and KIPPING, G.J. 1998a, The effects of line length and method of movement on patterns of reading from screen. *Visible Language*, **32**, 150–181.
- DYSON, M.C. and KIPPING, G.J. 1998b, Exploring the effect of layout on reading from screen. In R. D. Hersch, J. André and H. Brown (eds) *Electronic Documents, Artistic Imaging and Digital Typography* (Berlin: Springer-Verlag), pp. 294–304.
- FOSTER, J.J. 1970, A study of the legibility of one- and two-column layouts for BPS publications. *Bulletin of the British Psychological Society*, **23**, 113–114.
- GOULD, J.D., ALFARO, L., BARNES, V., FINN, R., GRISCHKOWSKY, N. and MINUTO, A. 1987, Reading is slower from CRT displays than from paper: attempts to isolate a single-variable explanation. *Human Factors*, **29**, 269–299.
- GOULD, J.D. and GRISCHKOWSKY, N. 1986, Does visual angle of a line of characters affect reading speed? *Human Factors*, **28**, 165–173.
- GRABINGER, R.S. 1993, Computer screen designs: viewer judgements. *Educational Technology and Research Development*, **41**, 35–73.
- GRABINGER, R.S. and AMEDEO, D. 1988, CRT text layout: perceptions of viewers. *Computers in Human Behavior*, **4**, 189–205.
- GRABINGER, R.S. and OSMAN-JOUCHOUX, R. 1996, Designing screens for learning. In H. van Oostendorp and S. de Mul (eds) *Cognitive Aspects of Electronic Text Processing* (Norwood, New Jersey: Ablex), pp.181–212.
- HARTLEY, J. 1994, *Designing Instructional Text* (London: Kogan Page).
- HARTLEY, J., BURNHILL, P. and FRASER, S. 1974, Typographical problems of journal design. *Applied Ergonomics*, **5**, 15–20.
- HORTON, W., TAYLOR, L., IGNACIO, A. and HOFT, N.L. 1996, *The Web Page Design Cookbook* (New York: John Wiley).
- HUEY, E.B. 1908, *The Psychology and Pedagogy of Reading* (New York: Macmillan; reprinted 1968 Cambridge, Massachusetts: MIT Press).

- JACKSON, M.D. and McCLELLAND, J.L. 1979, Processing determinants of reading speed. *Journal of Experimental Psychology: General*, **108**, 151–181.
- KAHN, P. and LENK, K. 1993, Typography for the computer screen: applying the lessons of print to electronic documents. *The Seybold Report on Desktop Publishing*, **7**, 3–16.
- KOLERS, P.A., DUCHNICKY, R.L. and FERGUSON, D.C. 1981, Eye movement measurement of readability of CRT displays. *Human Factors*, **23**, 517–527.
- KRUK, R.S. and MUTER, P. 1984, Reading of continuous text on video screens. *Human Factors*, **26**, 339–345.
- LUND, O. 1999, *Knowledge Construction in Typography: the Case of Legibility Research and the Legibility of Sans Serif Typefaces*, PhD thesis (Reading University, UK)
- MILLER, G.A. 1962, *Psychology, the Science of Mental Life* (New York: Harper & Row).
- MILLS, C.B. and WELDON, L.J. 1987, Reading text from computer screens. *ACM Computing Surveys*, **19**, 329–358.
- MORRISON, R.E. 1983, Retinal image size and the perceptual span in reading. In K. Rayner (ed) *Eye Movements in Reading: Perceptual and Language Processes* (New York: Academic Press), pp. 31–40.
- MORRISON, R.E. and RAYNER, K. 1981, Saccade size in reading depends upon character spaces and not visual angle. *Perception & Psychophysics*, **30**, 395–396.
- MUTER, P. 1996, Interface design and optimization of reading of continuous text. In H. van Oostendorp and S. de Mul (eds) *Cognitive Aspects of Electronic Text Processing* (Norwood, New Jersey: Ablex), pp. 161–180.
- MUTER, P. and MAURUTTO, P. 1991, Reading and skimming from computer screens and books: the paperless office revisited? *Behaviour & Information Technology*, **10**, 257–266.
- OBORNE, D.J. and HOLTON, D. 1988, Reading from screen versus paper: there is no difference. *International Journal of Man-Machine Studies*, **28**, 1–9.
- POULTON, E.C. 1959, Effects of printing types and formats on the comprehension of scientific journals. *Nature*, **184**, 1824–1825.
- RAYNER, K. and POLLATSEK, A. 1989, *The Psychology of Reading* (Hillsdale, New Jersey: Lawrence Erlbaum).
- RICHARDSON, J., DILLON, A. and McKNIGHT, C. 1989, The effect of window size on reading and manipulating electronic text. In E. D. Megaw (ed) *Contemporary Ergonomics 1989* (London: Taylor & Francis), pp. 474–479.
- SPENCER, H. 1968, *The Visible Word* (London: Royal College of Art).
- TINKER, M.A. 1963, *Legibility of Print* (Iowa: Iowa State University Press).
- TINKER, M.A. 1965, *Bases for Effective Reading* (Minneapolis: Lund Press).
- TWYMAN, M. 1981, Typography without words. *Visible Language*, **15**, 5–12.
- TWYMAN, M. 1982, The graphic presentation of language. *Information Design Journal*, **3**, 2–22.
- YOUNGMAN, M. and SCHARFF, L. 1998, Text width and margin width influences. <http://hubel.sfasu.edu/research/textmargin.html> (Presented at South Western Psychological Association).
- ZACHRISSON, B. 1965, *Studies in the Legibility of Printed Text* (Stockholm: Almqvist and Wiksell).