CAMBRIDGE
UNIVERSITY PRESS & ASSESSMENT

What impacts success in proofreading? A literature review of text feature effects

Research Report

Melissa Mouthaan

Sylvia Vitello

16 November 2022

Author contact details:

Sylvia Vitello
Assessment Research and Development,
Research Division
Shaftesbury Road
Cambridge
CB2 8EA
UK

sylvia.vitello@cambridge.org
https://www.cambridge.org/

As a department of the university, Cambridge University Press & Assessment is respected and trusted worldwide, managing three world-class examination boards, and maintaining the highest standards in educational assessment and learning. We are a not-for-profit organisation.

Cambridge University Press & Assessment is committed to making our documents accessible in accordance with the WCAG 2.1 Standard. We're always looking to improve the accessibility of our documents. If you find any problems or you think we're not meeting accessibility requirements, contact our team: Research Division
If you need this document in a different format contact us telling us your name, email address and requirements and we will respond within 15 working days.

How to cite this publication:

Mouthaan, M., & Vitello, S. (2022). *What impacts success in proofreading? A literature review of text feature effects.* Cambridge University Press & Assessment.

Acknowledgements

We wish to thank David Beauchamp for conducting the preliminary literature searches and review of the journal articles.

Contents

Executive Summary

Context

- Proofreading is an activity undertaken by professional proofreaders, but it is also a task performed in a variety of desk-based roles on an everyday or occasional basis by individuals who are not necessarily formally trained in proofreading. There is a clear organisational and professional interest in understanding the conditions that affect proofreaders' performance and experiences of this task.
- This report focuses on the potential impact of text features on proofreading. Our aim is to raise awareness of how design decisions for written materials may affect how well they are later proofread. Are there certain text features that are used in materials that are likely to be more 'high risk' in terms of proofreading?
- This report presents a literature review of the empirical literature that examined the relationship between text features and proofreading performance. The report also discusses various pieces of research from the readability literature, because there was a relative paucity of empirical studies that examined proofreading specifically.
- The review focused on three types of formatting and non-formatting text features:
    1. Line features - line length, line spacing and text alignment.
    2. Text formatting - text size, typefaces and fonts.
    3. Lexical features - word frequency, word predictability, homophones and word shape.

Findings

Overall

- The review revealed that the proofreading literature on text feature effects was very small and limited in terms of numbers, methodology and scope. This meant we had to draw more substantially on the readability literature. Together, this inevitably limits the conclusions that can be drawn about proofreading from the existing research literature, and highlights the need for further and more comprehensive research within the proofreading field.
- Despite the various caveats and limitations of the existing literature (on proofreading and readability), the review revealed patterns in the evidence. Most of the text features showed some relatively consistent evidence of affecting performance (line spacing, text alignment, text size, word frequency, predictability and word shape). However, there was less consistent evidence for text fonts and typefaces.
- Together, the evidence base enabled us to propose tentative recommendations for proofreading practice. However, all of these recommendations need further testing to determine how effective they could be at improving proofreading performance and the proofreading experience.

Line features

- The strength of the existing evidence on line features and proofreading outcomes was relatively weak. Research on readability offered the richest source of material in the general absence of proofreading-specific literature. However, findings were to some extent inconclusive, and in some cases contradictory.

- The review showed that the readability findings were conflicted on the subject of optimal line length, but there is evidence that 'extreme' line lengths (i.e. either long lines or very short ones) can disrupt ease of reading, such as when the reader (or proofreader) breaks their reading flow in order to locate the start of the next line.
- A clearer consensus emerged regarding line spacing in proofreading and readability research. There is evidence that wider line spacing (understood to be line spacing that is more than single-spaced) allows for more effective proofreading and reading, and that left-aligned text is preferable to justified text for reading on-screen.

Text formatting features

- The research on text size, typefaces and fonts in relation to proofreading performance and readability shows that there are useful insights, but also inconclusive evidence, in terms of impact on proofreading.
- Findings on text size are generally consistent. There appears to be sufficient evidence, and consensus, to conclude that larger characters allow better proofreading and reading performance (but not beyond a certain size).
- There is however limited evidence of a significant impact of font and typeface in both the proofreading and readability literature. There is evidence of a significant interaction between font and text size across several studies, while there is also evidence that different typefaces and fonts of a certain character size enable good performance when proofreading or reading.
- Some have argued that font, typeface, size and other formatting combinations should be assessed on a case-by-case basis, and generalisations about these individual text features on their own should be avoided.
- Finally, there is some consensus that variations in font type and typeface have a smaller impact than text size. Overall, proofreading and readability studies both provide indication that the impact of typeface and font type on proofreading is best considered holistically in relation to other factors, such as other font presentation factors (character size), and social factors (e.g. common preferences regarding font).

Lexical factors

- Although the research (in particular on proofreading) is limited in terms of scope, findings on the impact of lexical factors have been broadly consistent.
- The research on proofreading outcomes shows that some categories of errors such as homophone foils (e.g. *to-too; see-sea*) are more difficult to detect.
- There is also some further evidence from both bodies of literature that detection of certain errors is made more difficult by tendencies in the human visual and cognitive systems. For instance, readers respond differently to words of high/low frequency, with common, high-frequency words more likely to be skim-read or skipped by the reader. Reading time also decreases when words are predictable: that is, the reader is more likely in such a scenario to predict a word based on the preceding context rather than read it, which may in turn affect error detection.
- These studies allow us to pinpoint and summarise specific lexical features that are more likely to hinder error detection in proofreading. However, the application of these results for improved proofreading practice are discussed only in a very limited way in the literature – or are not treated at all.

Tentative recommendations for practice

- Based on the evidence as well as other researchers' recommendations within the literature, we note some suggestions for proofreading practice especially when the text is formatted with potentially challenging text features (and assuming it is not possible to alter these features)
- The proofreader may want to adopt other strategies, such as ensuring other distractions (e.g. background noise) are minimised, or double-checking text.
- For on-screen proofreading, proofreaders may be able to use features of a computer monitor to their advantage. This could include:
    - using appropriate software to mask parts of the text to focus on proofreading small sub-sections of the text at any one time.
    - limiting the amount of text visible on screen, although this warrants closer examination in further research to see if these are suitable for improving proofreading performance, or instead may create further issues.
    - using technology to assist in screening for circumstances where errors are more likely to occur, and draw the proofreader's attention to these. This could take the form of automatically screening for high frequency words or homophones in a document, and alerting the proofreader to potentially 'high risk' words or sections
- It may be useful for proofreaders and organisations employing proofreaders to be aware that personal preferences around certain text features (font, typefaces and text size) are not necessarily linked to actual proofreading performance.
    - For text features where there is less consistent evidence of effects on performance (e.g., font features), it may be more helpful to take readers' own preferences regarding into account.
- At an organisational level, organisations may want to check material intended for proofreading for unfavourable text features formatting, and use mitigation strategies for 'high risk' materials where formatting cannot be altered. This may for example include:
    - sending texts for proofreading to multiple proofreaders
    - sending shorter sections of text
    - allowing proofreaders more time to proofread text

Introduction

Proofreading is an activity undertaken within many organisations. It is primarily defined as a revision activity focused on identifying errors related to spelling, punctuation and grammar (SPaG) that deviate from a given standard (Bean & Bouffler, 1987; Porte, 2001). A broader definition is also often used that identifies proofreading as a task involving the checking of content, factuality, sentence structure, spelling and logic of writing (Chan & Ng, 2012). In this broader sense, and throughout this report, proofreading is understood as an activity that checks for two types of errors: spelling and related errors, but also semantic and contextual errors. Detecting this second type of error requires memory of previous parts of the text, interpretation of meaning, and integration of different parts of the text (Halin et al., 2014, p.70). Within organisations, in some contexts proofreading can encompass an even broader range of tasks, such as the checking of page numbers in a table of contents and whether this corresponds to page numbers within a document.

Research on proofreading has covered many aspects. One key topic of research on proofreading is the impact of 'text features', which refers to the presentation of text and the manipulation of typographical features. This has produced analyses of how text features influence proofreaders' error detection and pace of reading, but also the possible impact of text features on the proofreading experience itself, such as the extent to which text features influence whether a proofreading task is perceived as laborious or tiring by the proofreader. However there are many factors that can affect proofreading, and in the wider literature proofreading studies have also examined the effects on error detection in proofreading in relation to: proofreading medium (proofreading on-screen and on paper) (see Gould et al., 1987; Köpper et al., 2016); sound environments or background speech (see e.g. Jones et al., 1990; Venetjoki et al., 2006); and social factors, such as collaborative proofreading practices (Nihei et al., 2002). Research on cognitive factors also make up a distinct section of the proofreading literature, including for example studies that have examined research participants' age as a factor (Cohen, 1980; Levy et al., 1992).

Proofreading is of course an activity undertaken by professional proofreaders, but it is also a task performed in a variety of desk-based roles on an everyday or occasional basis by individuals who are not necessarily formally trained in proofreading. There is a clear organisational interest in creating conditions and implementing measures that allow for effective proofreading (where 'success' in proofreading is the successful detection of errors during proofreading). Understanding how different factors can impact proofreading performance has relevance for any organisation that has an interest in ensuring written materials are error-free. Similarly, it is relevant to understand how different factors impact on the proofreading *experience*. Understanding how text features may impact on proofreaders' subjective wellbeing when completing tasks is important, and steps taken to ensure a positive proofreading experience can contribute to proofreader retention. The proofreading 'experience' is understood in the literature, and throughout this report, as proofreaders' own perceptions of their proofreading performance, and their subjective experiences (e.g. satisfaction) in relation to proofreading under different conditions. This report's analysis considers both professional and occasional proofreaders when discussing findings, implications of findings, and when discussing proofreading strategies.

This report focuses on text features, summarising the research evidence of the impact of text features on error detection and the proofreading experience overall. This is one report of a two-part series of reports on proofreading. In our second report, we analysed the display mode used for proofreading, specifically comparing evidence of proofreading on screen and on paper (Vitello & Mouthaan, 2022). By understanding the impact of text features on proofreading, this report serves to raise awareness of the impact of different text features in the design of materials used, and if certain text features and formatting used in materials are likely to be more 'high risk' in terms of their potential impact on proofreading. In terms of implications for practice, we consider the following questions: Which elements of text formatting facilitate error detection in proofreading, and which may have a negative effect? How can we mitigate against any potential negative effects, both as organisations that employs proofreaders and as individual proofreaders? And, to what extent are possible actions, such as more proofreading time, additional proofreaders, and additional checks conducted by proofreaders, effective mitigation strategies?

This review on text features' impact on error detection in proofreading was conducted for the following purposes. Firstly, the purpose was to take stock of the research evidence on text features and answer the following research questions:

- What research has been done on proofreading in relation to text features?

- What is the strength of the existing evidence?

- What possible implications can be drawn to inform proofreading practice?

A further research question that arose over the course of this review was:

- To what extent does research into readability and text features have implications for proofreading?

An obvious potential use that can be drawn from research on text formatting is that it is informative for typesetting and the preparation of text for proofreading, in the sense that adjustments to features such as line length, text alignment and line spacing may increase the likelihood that errors are detected, but also improve the proofreading experience. Yet, this report primarily adopts a 'proofreader perspective' where the assumption is that proofreaders are often unable to manipulate typographical features or typesetting themselves. As a result, although the analysis and discussion are intended to raise awareness of the impact of various text features on proofreading, suggestions for the eventual preparation or manipulation of text, typographical and text features for optimal proofreading are not the focus in the discussion sections.

Finally, it is important to highlight that some of the reviewed literature on text features examines the impact of these features in relation to overall reading comprehension, rather than proofreading specifically. An initial review of the literature revealed a relative paucity of empirical studies that examined proofreading specifically, while in comparison text features and readability measures have been more extensively covered in the empirical literature on readability. For example, the literature on readability in assessment material discusses text-formatting features of paper and on-screen assessment, such as font type, size and style, as well as non-formatting features of text such as sentence and word length (Crisp et al., 2012; Macinska & Pastorino, 2020). Other areas of the readability literature have looked at the relationship between text features and visual stress, and have examined how the neural processing involved in reading text varies according to the characteristics of letter strokes associated with different fonts (Wilkins et al., 2020). Proofreading and readability studies have a different focus in terms of aspect of reading performance that they measure –

readability studies often examine reading comprehension and reading pace as measures of readability, while proofreading studies are more explicitly focused on error detection. We consider that the readability literature is nonetheless informative for proofreading, particularly in terms of overlapping discussions around reading comprehension and to some extent the impact of text features on the (proof)reading experience. The readability (or reading comprehension) literature therefore offers a useful reference point to understand potential impacts of text features on proofreading, but as a distinct field of inquiry the different nature of empirical studies on readability and proofreading is specified throughout this report's analysis.

The report is structured as follows. The text features covered in this report are divided into two sections: formatting and non-formatting features of text. Formatting features refer to presentation and the spatial organisation of text, where 'formatting' is understood to be the way text is presented or otherwise laid out on the page. Other text features discussed that are unrelated to text layout have been categorised as 'non-formatting text features', where those analysed in this review are lexical factors, i.e. relating to words and language. The report is divided thematically into sections. The conclusion of each section includes tentative suggestions for good practice in proofreading, and considerations that can be made in relation to text features and proofreading by proofreaders and organisations that employ them. Suggestions for practice are intended for both professional and occasional proofreaders.

Methodology

A traditional literature review was conducted by searching and accessing literature on proofreading outcomes in relation to text features. A large body of literature on readability was also accessed: given the common focus on text features, readability studies were determined early on in the project to have strong relevance for proofreading.

Literature searches were carried out in a number of stages. An initial thorough review was carried out in late 2020 using Google Scholar as a literature database, and subsequent literature searches were conducted using Google Scholar, ERIC and Scopus. No date filters were applied to the search. Some of the papers accessed were literature reviews on proofreading that featured relevant analyses of other studies (any use of secondary analysis is stipulated within this report). Further relevant studies were also identified and drawn from the reference lists of papers, and directly accessed and included in the analysis. Two papers that were generated from the search were subsequently excluded from the analysis because of a lack of relevance, or for featuring substantial errors including methodological weakness.

A total of 14 studies on text features and proofreading were analysed, where the majority of these were peer-reviewed, empirical studies. Other types of proofreading studies included were literature reviews, a commentary piece, and short research briefs. In addition, 14 studies on readability in relation to text features were included in the analysis, comprising primarily books and peer-reviewed studies (both empirical studies and literature reviews).

Searches were conducted using terms such as 'proofreading errors'; 'proofreading accuracy'; 'proofreading format of text', and containing at least one additional term such as 'font size'; 'typeface'; 'features'; 'format'. Subsequent searches were later also conducted using specific terms of interest such as 'line length' and 'line spacing'. While a systematic and comprehensive search was conducted to identify all relevant proofreading papers, a more selective approach was used to identify and include readability papers. The focus on proofreading also produced several studies on readability and many of these were included in the analysis, as well as any readability studies cited therein that were deemed relevant. A further search was conducted using the terms 'readability' or 'legibility' and containing other terms of interest such as 'font size'; 'typeface'; 'features'; 'format'. From the results that this search produced, any further studies that had not already been accessed and were recently published (2010-2021) were included. In addition, other research outputs published by Cambridge University Press & Assessment researchers in previous years on readability and accessibility were accessed and used in this report, while readability studies cited in these reports were also selectively accessed.

Text features: Formatting features

This section discusses formatting-related features of text. The following line features are discussed: line length (also often referred to as characters per line, or line width), line spacing, and text alignment. In addition, features related to the appearance of the text itself, especially text or font size, fonts, and typefaces, are also discussed. These features have been the subject of analysis in much of the empirical research on the impact of text appearance and formatting on proofreading and readability.

Line features

This section discusses the findings from proofreading and readability research on line length, line spacing, and text alignment.

Line length

A total of three articles were found that examined the effect of line length on proofreading, of which two are empirical studies and the third is a critical discussion piece. One of the empirical studies discusses proofreading in the readers' native language (in this case Chinese), while the two remaining studies were in the context of proofreading in a second language. These studies are presented directly below in reverse chronology. Then the section discusses studies on the effects of line length in the literature on readability of text that extends back to the early twentieth century. This body of research, much of it undertaken by psychologists and typographers, includes key studies undertaken in the 1940s-60s about line length and readability when reading on paper. The most common measurement of line length is number of characters per line (cpl), which is arguably the critical variable when considering line length (Dyson, 2004), although early studies have tended to measure line length in inches or millimetres.

Proofreading

Chan et al. (2014) conducted a study of on-screen proofreading performance in Chinese and compared three text formatting factors in their proofreading study: line length, line number (which they defined as the number of visible lines in the window view at any given time in the study), and line spacing. Specifically, they compared: (1) short, medium and long line lengths (which they defined as 26, 36 and 46 cpl respectively); (2) two, four and eight lines of text; and (3) single, 1.5 and double line spacing. 39 native-Chinese speaking undergraduates participated in the study, where each participant was tasked with proofreading 18 different passages under different line length, line number and line spacing conditions. The authors found that while there was a significant line number and line spacing effect on error detection (and that there was an interaction between the two), the effect of line length was non-significant (p. 526). However, they found that line length was negatively correlated with proofreading time and error detection in the sense that more scrolling led to faster proofreading, but poorer error detection performance. They therefore concluded that the display setting of medium line length (36 cpl) and 1.5 line spacing should be used. The authors found that this setting balanced proofreading time and error detection rate, and it improved performance for on-screen Chinese proofreading.

Porte (2001) looked at on-screen proofreading performance in a study involving 60 native Spanish speaking undergraduates enrolled at an English Foreign Language writing class at the University of Granada. The participants had all been classified as 'underachieving' in the writing class. Porte's error recognition task was set in four different line length conditions: 35, 45, 60 and 75 cpl respectively, while other text formatting factors remained constant (Times New Roman font, size 14 characters, and double spacing). Each participant was given a series of short texts to proofread in English, where a computer-generated frame highlighted four-line sections of the text a time, and masked all other surrounding lines. The participant was required to indicate any detected errors orally – only error recognition was required, and participants were not expected to correct the error. Porte's study found that a medium line length of 45 characters resulted in the best error detection performance, while a longer line length of 75 cpl resulted in the poorest error detection scores. Porte suggested that a medium line length setting ensured the total quantity of text visible on screen was limited, whereby error detection increased. This formed the basis of his conclusion that error recognition improves when there is 'less text available for visual sweeping' (p. 144). Interestingly, he found that error detection did not improve for the shortest line length (35 cpl) when compared to 45 cpl, suggesting that there is a point where reducing the amount of text visible no longer enhances the error recognition process. He argued that error detection among non-native speaking students could be improved if the students used scrolling or masking features of the computer monitor which would help divide the text up visually into more manageable units.

Cogie et al.'s (1999) article made a similar recommendation to Porte's study. The authors reflected on their experience of hosting a regular writing centre for ESL (English as a Second Language) students, and on strategies that students might apply to effectively check for errors in their own work. While the authors did not reference line length specifically, they discussed a number of self-editing strategies devised by the centre's staff, and applied by the ESL students in their cohort. In particular, Cogie et al. (1999) recommended using a ruler or piece of paper to cover text and proofread one line at a time, with the aim of focusing attention (p. 21). The authors described anecdotal evidence of instances where individual students applied this technique alongside other self-editing strategies, with the result that students' error detection when proofreading their own work improved substantially.

It is important to note that these studies provide only a limited scope to consider various line lengths and the impact on proofreading performance. The articles are few in number, implying a large research gap, and only two studies are empirical studies. In addition, Chan et al. (2014) examined proofreading performance in Chinese, while the other two studies examined error detection scores among non-native English speakers: a degree of caution is required when generalising these findings to understand the impact of text features in other languages (e.g. English), while professional proofreaders are likely to be either native speakers – or have a strong grasp – of the language they are proofreading in. In addition, a fairly short-to-moderate line length emerges as the optimal line length in both empirical studies, although only Porte's study tested line lengths over 45 cpl. For purposes of comparison, we note that the line length for standard paragraphs in our document is approximately 77 cpl (font: Arial 11pt). These limitations mean that a definitive claim on optimal line length in proofreading is difficult to make. In comparison, the literature on readability has tested longer line lengths and their suitability for different measures of readability, including reading comprehension.

Readability

Summarising both the early body of readability research and more recent studies, Nanavati and Bias (2005) noted that line length has been tested in relation to reading comprehension, reading speed, method of movement (e.g. scrolling or turning a page) and eye movements. Early research on line length appeared to reach a consensus that short-to-moderate line length is optimal for efficient reading on paper when compared to more extreme line lengths (i.e. very short or long lines) (Nanavati & Bias, 2005, pp. 124–125; Tinker, 1963). Several studies on readability in print for example argued that very short or very long lines interrupt the normal pattern of eye movement as the reader progresses through the text (Burt et al., 1955; Nanavati & Bias, 2005; Spencer, 1969). However, the definition of 'moderate' or 'medium' line length varies in both early studies of readability on paper, and in more recent literature that has examined readability on screen there is a lack of consensus in terms of optimal line length. [1] Some found that a medium line length of 55 cpl appeared preferable (Dyson & Haselgrove, 2001; McMullin et al., 2002) while other experiments have led to recommendations for a longer line length ranging between 85-100 characters per line (cpl) when reading on screen (Dyson & Kipping, 1998; Shaikh & Chaparro, 2005a, 2005b).

The different proofreading and readability studies and their findings on line length are summarised in Table 1 below.

---

1 It is relevant to note that research on optimal line length for reading on paper produced over the last century is not automatically relevant for on-screen reading. Nanavati and Bias (2005) argue that differences between reading hard copy and reading on a computer display should be taken into account, such as lighting source, glare potential, whether the text can be moved, and visual angle (pp. 122-123).

Table 1. Line length parameters as defined and evaluated in different studies on readability and proofreading.

Author(s), Year	Line length (cpl/inches/cm)				Recommendation on-screen/on-paper	Readability/proofreading factor measured in experiment
	Short	Short to moderate	Moderate to long	Long		
Ganayim and Ibrahim (2013)	-	7.37 cm	11.68 cm	24.62 cm*	On paper	Reading speed; Reading comprehension
Burt et al. (1955)	-	3.5 inches*	5.5 inches*	-	On paper	Reading speed; Reading comprehension
Chan et al. (2014)	26 cpl	36 cpl*	46	-	On screen	Error detection; Proofreading time
Dyson and Haselgrove (2001)	25 cpl	55 cpl*	-	100+ cpl	On screen	Reading speed; Reading comprehension
Dyson and Kipping (1998)	25cpl and 40 cpl	55 and 70 cpl	85 cpl*	100 cpl*	On screen	Reading speed; Reading comprehension; Perception of ease of reading
McMullin et al. (2002)	-	55 cpl*	-	115 cpl	On screen	Reading comprehension
Nanavati and Bias (2005)	-	55 cpl*	75 cpl*	-	Both	Reading speed; Reading comprehension
Porte (2001)	35 cpl	45 cpl*	60 cpl	75 cpl	On screen	Error detection
Shaikh and Chaparro (2005b)	35 cpl	55 cpl	75 cpl	95 cpl*	On screen	Reading speed; Reading comprehension; Satisfaction
Shaikh and Chaparro (2005a)	35 cpl*	55 cpl	75 cpl	95 cpl*	On screen	Reading speed; Reading comprehension; Satisfaction
Spencer (1969)	-	-	70 cpl*	-	On paper	-
Tinker (1963)	-	52 cpl*		-	On paper	Reading speed; Reading comprehension

*Line length(s) recommended by author(s) highlighted in blue.

The specific findings of readability studies are expanded on below, grouped according to line length recommendations (medium length; longer length; and mixed recommendations).

Medium line length

- Burt et al. (1955) found reading speeds and ease of reading on print was optimal at medium line lengths of 3.5-5.5 inches (between 89-140mm). Their recommendation was based on one legibility experiment of text produced in 10pt Times New Roman.
- Spencer (1969 cited in Nanavati and Bias 2005) recommended that line length when reading print should not exceed 70 cpl.
- Dyson and Haselgrove (2001) examined line length and reading comprehension in their study on readability on screen. They concluded that, based on the parameters of different line lengths included in their study, a medium line length of 55 cpl produced the best results in terms of comprehension. They found that long lines (which they defined as 100 cpl) could serve to disrupt readers' comprehension as the reader locates the beginning of a new line. However, the authors acknowledged that the line lengths tested for readability in their study were spread across a wide range, and that a better length might exist that was not tested in their experiments (for example, between 55-100 cpl).
- Research conducted by McMullin et al. (2002) tested line length and participants' reading comprehension when a text column was surrounded by white space, and when the text column was placed alongside an additional column of irrelevant text. They found that text presented in fairly narrow, single columns of 55 cpl, with surrounding white space, is preferable to long line lengths (115 cpl), for on-screen reading. While the authors did not find that line length affected reading comprehension in their experiments, they did find a small positive effect of having text surrounded by white space rather than irrelevant text, which they speculated acted as a distraction to the reader. The wide range between the two line lengths tested again does not rule out the possibility that a better line length might be found between those tested by the authors.

Longer line length

- Dyson and Kipping (1998) conducted two experiments where they examined line length in relation to reading rate, reading comprehension, and participants' own perceptions of ease of reading. They noted some inconsistencies between the experiments and reflected that there was no simple way to translate findings into guidelines for displaying text on screen. However, they suggested that a relatively long line length of 100 cpl allowed an efficient method of reading while minimising scrolling movements (which can also be disruptive). This recommendation thus differs from readability studies in print research that predicted a decrease in legibility over 70 cpl.
- In Arabic, Ganayim and Ibrahim (2013) examined the impact of line length on readability while comparing single and multiple columns of text. They found that reading comprehension of participants in their study was highest in conditions of a single column of text with long lines (where the column width they tested was 24.62cm) when compared against two or three columns of text (column width 11.68cm and 7.37cm respectively). They found that column-by-line-length layout had a significant influence on comprehension, but not on reading and comprehension speed (p. 330). The authors argued that visual complexity, such as connectivity, is an important determinant of reading in Arabic, and argued that this explained these results.

Mixed recommendations

- In experiments conducted by Shaikh and Chaparro (2005a, 2005b), the authors examined readability in relation to different line lengths, and – in one series of experiments – different passage types. They determined from their first experiments that line length had an impact on reading efficiency (computed by multiplying reading speed by reading comprehension) and that longer lines of 95 cpl were more efficient than lines of 35cpl. The authors also looked at effect of line length on satisfaction and preferences, and found no conclusive results: there was high variation in preferences, and, while some participants perceived themselves to be reading faster at 35cpl, this line length setting actually resulted in the slowest reading speed (Shaikh & Chaparro, 2005a, 2005b). In a further study comparing two passage types (longer, narrative passages and short, online news articles) and line length, Shaikh and Chaparro found that line lengths of 95 cpl resulted in the fastest reading rate for both passage types, but that short lines of 35 cpl resulted in the highest comprehension score for narrative passages with longer sections of text (Shaikh & Chaparro, 2005a).

Line spacing and text alignment

Proofreading

This section discusses line spacing and text alignment. Not many studies have looked at text alignment – those that have done have examined text alignment at the same time as line spacing. Only one proofreading study was found and analysed that examined line spacing and proofreading performance. No studies on text alignment and proofreading were found. In the previous section on line length, one proofreading study was described (Chan et al., 2014) that also examined various line spacing settings. The authors tested three line spacing settings: single (1), one-and-a-half (1.5), and double spacing, and recommended the use of 1.5 line spacing, although noting that the error detection rates between 1.5 and double spacing among their research participants was not statistically significant. They theorised that single spacing is relatively crowded resulting in greater interference from surrounding characters, making it difficult to detect errors.

Readability

Readability studies have examined line spacing and text alignment mainly in the context of on-screen reading of webpages. These studies have generally found that (1) additional line spacing is preferable to single line spacing, and (2) left-aligned text is usually preferable to justified text. For example:

- Lynch and Horton (1999, cited in Nanavati & Bias, 2005) found that added line spacing in text displayed on webpages improves legibility for longer lines of text.
- Similarly, Hill (2001, cited in Nanavati & Bias, 2005) found that lines with more than 9-10 words benefit from additional line spacing for readability purposes.
- Grabinger and Osman-Jouchoux (1996) summarised findings from the literature on print and online screen reading in relation to typography practices and found much of the research to be contradictory. However, they found some consensus that left-aligned text is preferable to fully justified text, which the authors perceived may affect reading comprehension because of hyphenation and uneven breaks in sentences. They also recommended the use of double spacing for longer lines of 60cpl, observing that longer lines made it harder for the reader to maintain position on that line (p. 196).

- Ling and van Schaik (2007) compared text alignment (justified and left-aligned text) and three levels of line spacing (single spacing, space-and-a-half, and double spacing) in relation to speed and accuracy of visual search of webpages and reading. They found that wider line spacing produced better reading accuracy than single line spacing, with double spacing being optimal. They also found that justified text was more difficult to search through, although increasing line spacing to double-spacing compensated for this. In addition, participants' subjective views did not correspond to performance: while participants preferred justified text, left-aligned text was found to produce better performance.

Discussion

- What research has been done on proofreading in relation to text features (in this case, line features)?

This section found and reviewed three studies (of which two were empirical studies) on proofreading in relation to the line features analysed. As noted earlier, the small number of studies indicates a large research gap in this area.

- What is the strength of the existing evidence? To what extent does research into readability and text features have implications for proofreading?

The strength of the existing evidence on line features and proofreading outcomes is relatively weak. Research on readability offered the richest source of material in the general absence of proofreading-specific literature. However, findings are to some extent inconclusive, and in some cases contradictory: researchers report that it is not easy to simply translate findings into guidelines for displaying text on screen, and have reported inconsistencies when examining findings from their own experiments. The proofreading studies offer some evidence that a short-to-moderate line length is preferable (36-46 cpl) and similarly there is some agreement among several readability studies that medium or moderate line lengths of around 55 cpl facilitate reading comprehension. However, there is no strong consensus regarding optimal line length recommendations for readability, and some of these studies have suggested that line length has a small or negligible effect on readability measures. There is also evidence that wider line spacing (understood to be line spacing that is more than single-spaced) allows for more effective proofreading and reading, and that left-aligned text is preferable to justified text for reading on-screen. The findings from readability research may serve as a good starting point for further studies exploring optimal line formatting for proofreading.

It is important to acknowledge that the methodological features of the studies cited above sometimes differ in ways that make generalisations difficult, and limit the ability to draw implications for real-world proofreading practice. These differences include different text passage types tested in reading comprehension experiments (e.g. narrative text passages, or web-pages); samples of research participants that include speakers who do not have strong ability in the language they are proofreading in; or experiments that simultaneously test the interference of other text factors, such as columns of irrelevant text presented to the reader alongside the text to be examined. Overall, this suggests that there is a need for further empirical research into both proofreading performance and readability in relation to line length, line spacing, and text alignment.

- What possible implications can be drawn to inform proofreading practice?

There are limitations on the implications that can be drawn for proofreading due to the limited available literature on proofreading performance and line features. The literature on line features in relation to readability is more extensive, and arguably offers some indication of appropriate line length, line spacing and text alignment for successful proofreading. While the analysis showed that the readability findings are conflicted on the subject of optimal line length, a clearer consensus emerged regarding line spacing in proofreading and readability research.

The introduction of this report stated that findings on line length, line spacing and text alignment have more obvious implications for typesetting rather than adaptations that proofreaders can make in their practice. While acknowledging these constraints, a number of tentative implications for proofreaders, organisations employing and/or managing proofreaders and for researchers are stipulated below.

- Given the relative paucity of empirical studies of the impact of line features on proofreading, the strongest implication is that there is a need for much more research in this area.
- There is evidence that 'extreme' line lengths (i.e. either long lines or very short ones) can disrupt ease of reading, such as when the reader (or proofreader) breaks their reading flow in order to locate the start of the next line. There is also some evidence that single spacing negatively affects error detection in proofreading and general reading comprehension. When reading text formatted in this way (and assuming it is not possible to alter line length or spacing settings), the proofreader may want to adopt other strategies, such as ensuring other distractions (e.g. background noise) are minimised, or double-checking text. At an organisational level, organisations may want to check material intended for proofreading for unfavourable line formatting, and use mitigation strategies for 'high risk' materials where formatting cannot be altered. This may for example include sending texts for proofreading to multiple proofreaders.
- Detecting the start of a new line has been shown to be more challenging with text displayed in longer line lengths. Some strategies have been suggested in the literature to facilitate detecting the start of a new line when proofreading, such as limiting the amount of text visible using an aid (e.g. a piece of paper). These strategies might be useful for proofreading practice, although they have largely not been the focus of controlled experiments in the literature.
- For on-screen proofreading, proofreaders may be able to use features of a computer monitor to their advantage. This could include using appropriate software to mask parts of the text to focus on proofreading small sub-sections of the text at any one time. Other strategies for limiting the amount of text visible on screen, such as scrolling, the 'zoom' function and the Immersive Reader feature in Microsoft 365, warrant closer examination in further research to see if these are suitable for improving proofreading performance, or instead may create further issues.

Text size, typefaces and fonts

A large section of the literature has examined text size, typefaces, and different font types, and how these interact at various levels with proofreading outcomes and readability. Proofreading studies make up a relatively small proportion of this body of literature on these text features. In the first instance, text size is discussed in relation to findings in proofreading and readability studies. Secondly, the distinction between ‘serif’ and ‘sans serif’ typefaces is defined and an overview is given of the debate on typeface and font implications for proofreading performance and general readability.

Text size

In both proofreading and readability literature, studies have widely found that legibility decreases when text size decreases. This is argued to be because visual acuity is limited for smaller text sizes, whereas a larger text size produces a larger retinal image (Dyson, 2004; Piepenbrock et al., 2014).

Proofreading

Three empirical studies that examined text size and proofreading performance in error detection were found and analysed, and are discussed in this section.

Piepenbrock et al. (2014) examined the impact of text size on proofreading performance in the context of their wider study on polarity. The authors tested their theory that any positive polarity[2] advantage in reading should become greater when character size decreases, and conducted a study with 165 native German-speaking volunteers as research participants[3]. Participants were placed in a dark room without light sources other than the computer display for the proofreading task, and were tasked with reading 40 texts of 250 words each. Participants were asked to read out loud any detected errors in the text, and were given 50 seconds per text to identify errors, which prior testing had shown was generally insufficient time to read the entire text. Piepenbrock et al. tested four text size conditions: 8pt, 10pt, 12pt and 14pt text in Helvetica font. The authors found that more errors were detected with increasing character size, while dark characters on a light background – positive polarity – led to better error detection for all text size conditions. Participants also read more words as text size increased, and in the positive polarity condition for all text size conditions tested in the experiment. A post-task questionnaire showed that participants themselves perceived no significant differences between the polarity displays in terms of text readability, but the questionnaire did not appear to examine participants’ subjective experiences of different text size conditions. Overall, Piepenbrock et al.’s research supports the general recommendation to present text in positive polarity, but particularly so when reading small characters. Their findings are particularly relevant when proofreading on devices that display text in small characters. While this generally refers to handheld devices such as smartphones, screen size may also be limited in other devices.

Chan and Ng (2012) conducted an on-screen experiment with 27 native Chinese undergraduates to examine the influence of font size, typeface, text direction and copy placement on proofreading speed, accuracy and subjective preferences in a proofreading task. In the task, participants were required to proofread different passages of text under 16 different text formatting combinations. Under each passage, participants were presented with buttons to indicate types of errors in the text (extra word, missing word, wrong word…).

2 Positive polarity refers to dark characters on a light background; negative polarity refers to light characters on a dark background.

3 It is likely that the proofreading task was in German, but this is not directly confirmed in the article.

The authors tested a ‘small’ font size (defined as 10pt) and ‘large’ font size (14pt) in a Chinese serif and sans-serif font (Ming Liu and Jheng Hei respectively). In general, they found that participants proofread faster in 14pt fonts, but that all four of the main factors tested did not have any significant effect on proofreading accuracy (for both the overall and individual error type detection rate). However, the preferences data revealed that participants expressed a significantly greater preference for 14pt characters, while proofreading time per character correlated significantly with the error detection rate: in other words, proofreading time per character was longer when the task was perceived by the participant to be less comfortable and more fatiguing. These findings led the authors of the study to recommend 14pt characters for Chinese proofreading.

Tullis et al. (1995) conducted a small-scale study with 15 volunteers where they examined the impact of different combinations of font and text size combinations on proofreading accuracy, proofreading time, and subjective preferences. The text sizes tested by the authors were relatively small when compared to the larger sizes tested in the two previously cited studies. A total of twelve different font and size combinations (character sizes ranging from 6.0 to 9.75pts) were tested. Participants were instructed to read a paragraph of text in each font/size combination, where each paragraph contained between 1-5 typographical errors, where the error(s) would consist of one randomly selected letter being replaced by a different, incorrect letter. The authors found that there was a significant interaction of font and text size on proofreading time, accuracy, and preferences. Participants were most successful in detecting errors in several of the larger font conditions included in the experiment. Participants also preferred 9.75pt text size conditions (specifically, Arial 9.75 and MS Sans Serif 9.75). The findings led the authors to recommend avoiding smaller fonts (such as Arial 7.5), and to instead select a font and size combination that optimises subjective preference without sacrificing speed and accuracy (such as Arial 9.75 or MS Sans Serif 9.75) when preparing text for proofreading.

Readability

Readability studies have also shown that text between certain sizes is associated with better readability. For example, a character size ranging between 9 and 12 point is widely regarded as the most legible for text read at normal reading distances (Lonsdale, 2014), while Bernard et al. (2003) found that reading accuracy is not significantly different between 10 and 12 point character size. Early research on readability on print in relation to text size also focused on lowercase and capital letters. These studies’ findings showed that lowercase letters were less legible than individual capital letters (Arps et al., 1969; Sheedy et al., 2005; Tinker, 1963). They attributed the greater legibility of capital letters to their larger size, and the fact that lowercase letters more often have only slight differences between them (e.g. *e*, *o* and *c*). Lowercase words, however, were found by Tinker and Paterson (1939) to be more legible than uppercase words when text size was equal. The authors speculated that this finding could result from the ‘less characteristic’ word forms inherent to text presented in capital letters, while lowercase words tend to have shape information in the form of a pattern of ascending or descending word height (Sheedy et al., 2005; Tinker & Paterson, 1939). This may make lowercase words more legible, although, as we note later in this report, there is some evidence that errors that do not alter a word’s expected shape are more likely to pass undetected during proofreading. See Monk and Hulme (1983) study discussed later in this report.

Typefaces and fonts

A number of proofreading and readability studies have examined typefaces and different fonts. In terms of letter styles, a key distinction is often made between serif and sans serif typefaces. Serif typefaces have cross-strokes that project from the main stroke of a letter, and have often been considered more visually appealing and readable. Sans serifs do not have this characteristic. It has been suggested that serifs 'generally aid in the process of reading by helping to distinguish each individual letter in a word' (Bernard et al., 2003, p.824).[4] Sans serif typefaces have become widely used in everyday life despite strong opposition from 'traditionalist designers and writers' (Beier, 2009, p.119).

This section discusses five studies on proofreading that have examined typeface or font type as a main factor (three studies), or as a possible influence (two studies).

Proofreading

Some early studies on proofreading tentatively attributed their findings on proofreading accuracy and error detection to font type. For example, Creed et al. (1987) compared error detection performance among their participants and noted that those proofreading on a Visual Display Unit performed poorly when compared to those proofreading on print, and suggested this could be due to differences in font appearance.[5] Similarly, Healy et al. (1983) attributed differences in the pattern of proofreading errors in their experiments to differences in font type used in their study. Although not conclusive, this indicates that the impact of font type has emerged as a consideration in proofreading research since the 1980s. The following studies have made a concerted effort to examine typeface and font effects.

Halin et al. (2014) carried out two on-paper proofreading experiments with students at the University of Gävle where the purpose of the study was to investigate the interactions between font type and presence/absence of background noise, and their impacts on reading speed and error detection. Participants' subjective perceptions of task difficulty were also recorded. 29 and 31 native-language Swedish students participated in each experiment respectively and were presented with four passages of text in Swedish of approximately 1020 words, containing both semantic/contextual errors and spelling errors (spelling error consisted of either missing or substituted letters). In the first experiment, each individual text was presented in either Times New Roman (a familiar and therefore 'normal' font status) or Haettenschweiler font (unfamiliar 'altered' font status), and sound conditions alternated between task-irrelevant speech played during the proofreading task, or silence. In the second experiment, conditions were identical except that all texts were written in Times New Roman, and the passages written in Haettenschweiler in the first experiment were masked by visual noise superimposed in greyscale over the page.

The authors found that proofreading in Haettenschweiler font was rated as more difficult and taxing by participants. However, they also found that, although background speech negatively affected the detection of semantic/contextual errors in the normal font status, this effect was not reproduced in the altered font status. They argued this was due to the

4 Fonts are either serif or sans serif. Times New Roman is an example of a serif typeface, while Arial is a common sans serif alternative.

5 Early computers did not have the same range of fonts, and so the fonts could not be identical to what could be printed.

‘shielding’ effect of an unfamiliar font, which demanded greater task engagement and concentration from participants. In the second experiment, they found that visual noise made the detection of spelling errors more difficult, whereas background speech had the opposite effect. The authors concluded that both text status alterations (visual noise and Haettenschweiler font) made spelling error detection more difficult, but visual noise was more detrimental than the unfamiliar font.

Chan and Ng (2012) examined proofreading performance in Chinese in relation to typeface and font size.[6] Similarly to Tullis et al., the authors found a significant interaction between the two (in this case, on speed of proofreading). Proofreading in 10pt font condition was found to be faster with serif font Ming Liu than sans serif font Jheng Hei (p. 1327). However, proofreading with the larger 14pt character size was significantly faster with Jheng Hei than Ming Liu. The authors argued that serifs thus appear to assist in distinguishing small, but not large, characters in Chinese proofreading. However, proofreading accuracy of participants in their study was found to be unaffected by typeface and font size. They argued their results show that proofreading accuracy can be obtained at the cost of a speed-accuracy trade-off: to ensure a certain level of proofreading accuracy, they recommended that proofreaders be given sufficient time in real-life proofreading tasks.

Tullis et al. (1995) examined the impact of font type on accuracy and preference in on-screen proofreading. They created and compared twelve font and size combinations from four font types (Small Font, Arial, MS Sans Serif and MS Serif). Their selection comprised both serif and sans serif typefaces.[7] They found that larger Arial and MS Sans Serif fonts (both sans serif) produced a better accuracy performance than other fonts tested. They also noted that participants generally preferred two of the sans serif fonts (MS Sans Serif and Arial) in combination with character size 9.75pt. Both accuracy and preference data indicated a significant main effect of font and size. The authors did not make any general recommendations regarding using a specific typeface for proofreading, and concluded that any of the fonts they tested allowed for good results in proofreading speed and accuracy at 8.25, 9.0 or 9.75pts, with the exception of MS Serif (8.25pt).

Readability

Some early studies on readability have argued that serif typefaces have better readability in print when compared to sans serif typefaces (Burt et al., 1955 and Labuz, 1988 cited in Bernard et al., 2003), while other studies have found there to be no difference between serif and sans serif typefaces for readability on print, or only a very limited impact (e.g. Tinker, 1963; for a review of other early studies see Lonsdale, 2014). In addition, some have argued that any possible added readability of serif typefaces does not extend to on-screen reading, with the positive effect of the typeface for readability being either reduced or eliminated altogether (Horton, 1989, cited in Bernard et al., 2003). Design guidelines for websites have led to sans serif fonts being more commonly used for online text, and more recent studies of readability on screen have contested the claim of superior readability of serif typefaces, finding no effect of font type or typeface on readability, or on visual search and information retrieval performance when reading online (Bernard et al., 2003; Ling & Van Schaik, 2006).

6 See previous section on ‘text size’ for further details on the study conducted by Chan and Ng (2011).

7 See previous section on ‘text size’ for further details on the study conducted by Tullis et al. (1995).

These studies demonstrated a subjective preference for sans serif fonts and, in the absence of a measurable impact of typeface on reading accuracy, recommend taking readers' own preferences into account. Crouwel (2001) theorised that the growth of a wide range of media has led reading habits to change significantly, and suggested that the same experiments conducted on typefaces and legibility in the present would perhaps show no difference between typefaces tested (Crouwel, 2001, cited in Beier, 2009).

The effect of font and typeface – and interactions with external factors – on legibility or reading performance seems to be complex, making generalisations difficult. Legibility of typefaces is therefore still a topic of discussion. The next section briefly discusses the findings of three of the most recent readability studies on font type and typefaces, where these studies examined font familiarity, the legibility of different fonts and typefaces, and perceptions of text legibility based on font differences.

- Beier (2009) carried out a series of experiments, with participants in each study conducting reading tasks with fonts that were both familiar and unfamiliar to participants. The author's aim was to examine the impact of font familiarity on the reader, where the experiments measured both reading speed and the reader's own perception of their pace of reading. 34, 41 and 60 students from the Royal College of Art and Imperial College took part in each experiment. The author found that the two familiar fonts compared favourably in terms of reading speed to the unfamiliar font – readers also perceived their performance to be better in the familiar font conditions. However, the author also found that the font familiarity effect on reading speed was short-lived: after being exposed to the unfamiliar font, readers quickly adjusted themselves and attained similar reading speeds to their speeds with familiar fonts.
- Sheedy et al. (2005) examined the effects of font and display parameters on text legibility in a series of experiments. In each experiment, 25-30 participants were recruited to take part from Ohio State University. In one experiment, they tested the readability of three serif fonts (Georgia, Times New Roman and Plantin) against three sans serif fonts (Verdana, Arial and Franklin) by measuring participants' visual acuity. Their findings showed that font type had a significant main effect: the two most legible fonts were sans serif, however the remaining sans serif font (Franklin) was the least legible. The authors concluded that a generalised statement about typeface legibility was not possible, and that future research should seek instead to determine the legibility of fonts on a case-by-case basis.
- Bernard et al. (2003) recruited 35 native-English speaking undergraduate and graduate students from a mid-western city in the United States to take part in their study on the on-screen readability of several different typeface, size and format combinations. The authors used a modified proofreading task to assess reading accuracy and speed in different font and text size conditions: participants were required to read eight passages of text and correctly identify substitution words (15 in each passage). Data on participants' perceptions of text readability was collected through a questionnaire after each passage of text, and a ranking of text combinations for preference after the experiment. Analysing the percentage of detected substituted words for each combination, the study did not reveal any significant differences for the different typeface, size and format combinations. However, the authors noted that participants slowed their pace of reading for less-readable passages (which they defined as text in smaller size), and argued this possibly enabled participants to achieve the same level of

reading accuracy. The study also noted a significantly greater subjective readability (participants' own perceptions of text legibility) between the conditions, with participants generally preferring Arial to Times text at both 10- and 12pt text size. The authors speculated that sans serif typefaces may be perceived as more readable on-screen, even in the absence of objective differences in readability between the two typefaces.

Other forms of text manipulation have also been examined in relation to font, and have formed part of the early literature on readability in printed form in relation to text features. For example, bold font and increased stroke width were shown to improve legibility of fonts (Luckiesh & Moss, 1940), and text in italics has been argued to reduce reading speed, considered to be a measure of legibility (Tinker, 1963). Given the absence of proofreading studies on these features, this section of the readability literature is considered to extend outside the scope of this report and is therefore not examined in greater depth.

Screen display characteristics also interact with, and therefore affect, the appearance of typefaces. One such display characteristic is the 'aliasing' (or 'staircasing') effect associated with characters on screens that can make letters look jagged on screens with sub-optimal resolution (Bernard et al., 2003). Anti-aliasing is the attempt to counteract this effect through techniques designed to make text more readable (for visual examples of anti-aliased text and other font-smoothing techniques, see 'Font Smoothing Explained' – Szafranek (2009) and Bernard et al. (2003, p.825). These techniques include font 'smoothing' within graphical images (e.g. JPEG files), or converting to Adobe Portable Document Format (PDF) file format. Anti-aliasing adds shades of darker contrast as a way of reducing contrast differences between the background and the letters. However, anti-aliasing techniques may work less well for certain fonts, typefaces and text size. Anti-aliasing may therefore render the text *less* readable if the font smoothing process blurs letterforms. For example, in their experiments Sheedy et al. (2005) tested different font and font-smoothing combinations, and found an interaction effect between font smoothing and font type: while 'ClearType' (subpixel font-smoothing) enhanced legibility of Georgia, Plantin and Franklin, it appeared to reduce legibility of Times New Roman and Arial. Similarly, Bernard et al. (2003) found in their study that anti-aliased 10pt Arial text was read more slowly than all other 12-pt text and dot-matrix text tested in their experiment.

Discussion

What research has been done on proofreading in relation to text features (in this case, text size, typefaces and fonts?)

This section found and provided an in-depth review of four empirical studies that explicitly examined the impact of text size, typeface and/or font on proofreading. At least two of these studies were conducted in languages other than English.

What is the strength of the existing evidence? To what extent does research into readability and text features have implications for proofreading?

The research on text size, typefaces and fonts in relation to proofreading performance and readability shows that there are useful insights, but also inconclusive evidence, in terms of impact on proofreading.

Findings on text size are generally consistent. Two proofreading studies determined that increasing character size was associated with greater success in error detection. One proofreading study in Chinese found that proofreading accuracy did not vary between 10pt and 14pt text, but that proofreading speed was faster in the 14pt character size condition. The readability literature similarly indicates a consensus that larger characters of around 12-pt text is preferable to smaller text in terms of common readability measures. However, there is some evidence of a limit on reading accuracy gains made beyond a certain increase in character size. In addition, both proofreading and readability studies have frequently collected data on participants' own preferences regarding text size, and concluded that participants preferred larger character sizes. There appears to be sufficient evidence, and consensus, to conclude that larger characters allow better proofreading and reading performance (but not beyond a certain size).

There is however limited evidence of a significant impact of font and typeface in both the proofreading and readability literature. Regarding the impact of unfamiliar fonts, one proofreading study noted that the use of unfamiliar font was less detrimental to error detection than other distractions (such as visual noise), while a readability study concluded that 'font familiarity' can be rapidly gained with exposure. There is evidence of a significant interaction between font and text size across several studies, while there is also evidence that different typefaces and fonts of a certain character size enable good performance when proofreading or reading. A few early studies recommended the use of serifs for readability, but this recommendation has been contested in later studies that have noted a small or negligible effect of font and typeface on reading accuracy and speed when reading on-screen. In a few cases, the different experiments produced some evidence of better proofreading and reading accuracy when reading sans serifs, but the data is too limited to be able to make a generalised statement about typefaces – and this is reflected in the studies' conclusions. There was also some evidence of subjective preferences for sans serif fonts.

The limited evidence of impact of typeface and font type on objective readability measures (speed and accuracy) has led some authors to recommend taking readers' own preferences regarding font into account. In addition, some have argued that font, typeface, size and anti-aliasing combinations should be assessed on a case-by-case basis, and generalisations about these individual text features on their own should be avoided. This recommendation may be useful to inform further research into the impact of font, typefaces and text size on proofreading.

Finally, there is some consensus that variations in font type and typeface have a smaller impact than text size. Overall, proofreading and readability studies both provide indication that the impact of typeface and font type on proofreading is best considered holistically in relation to other factors, such as other font presentation factors (character size), and social factors (e.g. common preferences regarding font).

What possible implications can be drawn to inform proofreading practice?

- Experiments examining text size, fonts and typefaces, font-smoothing, and interaction with display medium (e.g. screen) may produce different results if the same experiments were conducted today. For instance, many of the early claims made against sans serif typefaces, while initially extensive, may have limited applicability today as the use of sans serifs has become commonplace, and display technology is constantly evolving. There is therefore a need for ongoing, up-to-date research into the effect of these and other text features under conditions that reflect those of typical proofreading tasks in a contemporary setting.

- The evidence on serif and sans serif typefaces is mixed, and there is no clear indication that one typeface is better suited for proofreading than another. In addition, there are few empirical studies on this topic that have been conducted in recent years. While there is no strong evidence that proofreading in a specific font or typeface has a significant effect on error detection, there is a need for additional research on specific font-text size-typeface combinations.
- There is evidence that small text size is a factor that can impact proofreading, and interacts with other text features discussed in this section. Organisations may want to adopt certain strategies when requesting proofreading of text presented in smaller characters, and be aware that proofreading smaller text has been reported to be more tiring and laborious for the proofreader. Strategies may include allowing proofreaders more time to proofread text in small character sizes, or sending shorter sections of text to multiple proofreaders.
- It may be useful for proofreaders to be aware that personal preferences around font and typefaces are not necessarily linked to actual proofreading performance. Some studies have observed a divergence between objective readability (measured as reading or proofreading accuracy, or speed of reading) and participants' own preferences regarding fonts, typefaces and text size. As such, an individual's own perceptions of legibility and proofreading performance based on font or typeface do not necessarily correspond to actual error detection or pace of reading. In addition, while there is some indication that proofreaders and readers prefer, read faster, and are more successful at detecting errors in text written in a familiar font, there is also evidence that familiarity with a font can be gained fairly quickly.

Text features: Non-formatting features

This section discusses lexical factors as an area of text features that are unrelated to formatting but that may affect error detection in proofreading. Lexical factors are understood to be factors relating to words and language itself.

Lexical factors

Lexical factors are a key aspect that studies on both proofreading and readability have examined. Proofreading studies of lexical factors have generally examined word frequency, word predictability and eye movements. In assessment research, readability has been examined in relation to linguistic features, accessibility and assessment materials (e.g., see Beauchamp & Constantinou, 2020). Much of the literature has focused on inherent tendencies in the visual system, typically looking at aspects of reading and proofreading behaviour (particularly gaze fixation), in response to lexical factors such as low/high frequency words (defined as words that appear frequently or infrequently in a language) and word predictability (whether the reader is more likely to produce the next word in a sentence themselves based on the preceding context). In both proofreading and readability research, there is an early body of research on this topic (dating from 1975-1999) but little recent research.

Proofreading

Three empirical proofreading studies were found that looked at impacts of lexical features on proofreading, and are reviewed in this section.

Daneman and Stainton (1991) conducted a study on proofreading and homophones[8] where they recruited undergraduates at the University of Toronto to participate in a series of experiments. Between 32 and 64 participants were recruited for each experiment, and the authors first tested the reading comprehension levels of participants using the Nelson-Denny Comprehension test to ensure some variation within the sample. Participants were given a 1,100 word text to proofread that contained both homophone and non-homophone errors (e.g. *board-bored-beard*). The experiments were based on prior research that demonstrated that error detection was lower for incorrect words that nonetheless sounded correct, which they referred to as a 'homophone effect'. It is argued that homophone errors occur because readers activate phonological codes during reading, whereby incorrect words that sound the same as correct words are more likely to pass undetected (Daneman & Stainton, 1991, p.620). The authors reproduced this finding in their experiments and furthermore found that, even when participants were first given an error-free version of the text to familiarise themselves, the homophone-error effect persisted. The authors' explanation for this latter result was that participants read the text less carefully on a second reading than on the first. In addition, the study found that participants who had scored lower on reading comprehension nonetheless were not disproportionately affected by the phonological properties of the error words when compared to above-average readers in the sample. The authors argued that only a severe reading disability might produce a 'phonological deficit' that would lead to poorer performance in detecting this kind of error (p. 623).

Jared et al. (1999) used linguistics theory in their study on error detection and word substitution errors using homophones. They examined the use of different 'routes' in the activation of word meaning during a proofreading task, and the implications for proofreading accuracy. The authors conducted a series of experiments with introductory psychology students at different North American universities. Between 24 and 80 students took part in each experiment. Similarly to Daneman and Stainton's research, in some of these experiments participants were selected to represent a range of reading skills (defined as scoring above or below defined percentiles of the Nelson-Denny Comprehension test). Jared et al. applied errors to text that reflected incorrect homophones (e.g. *see-sea;* or *alter-altar*), where (i) low/high word frequency of the correct homophone and (ii) predictability of the correct homophone were factors manipulated in the study. Their study of word predictability in relation to error detection and homophone errors found that there was 'clear evidence' that predictability of the correct target word exacerbated the homophone effect on error detection, but particularly so for high-frequency words (p. 258). They also found that good readers performed better in terms of detecting pseudowords (errors in the form of non-words) as part of the experiments, and read faster (had a shorter gaze duration) than poor readers. Citing Daneman and Stainton's earlier proofreading experiments, their own experiments, and other studies on eye movement tracking in reading, Jared et al.'s study provided further support for the claim that 'homophone foils' are generally more difficult to identify than spelling errors. The authors also reiterated that further research should similarly control for varying reading skill levels.

8 A homophone is a word that is pronounced like another word but is different in meaning or spelling.

Monk and Hulme (1983) examined whether changes made to a word's shape (where they examined height) had a negative impact on proofreading accuracy. They recruited undergraduate students and sixth formers applying to the University of York as research participants for two experiments, with 60 participants taking part in each study. Unlike Daneman and Stainton and Jared et al.'s studies, they did not control for different reading skills among their participants. Their hypothesis was that misspellings that change a word's shape (i.e. when the height of a letter substituted in the word is different to the original) are more likely to be detected in proofreading than misspellings where the word's essential shape remains the same. The experiments also controlled for word frequency. Their research found a strong 'word shape' effect, whereby misspellings that maintained a word's shape were less noticeable to research participants than those that did not. They also found no evidence that this 'shape effect' depended on word frequency or word length. Monk and Hulme argued that the results of their study were unsurprising, given that the visual system is far more effective in detecting *patterns*, such as word shape, than detecting absolute quantities, such as word length. This research might help to explain why some spelling errors are not detected during proofreading.

Readability

Staub and Rayner (2007) provided a review of the early readability literature on reading behaviour in response to lexical factors. They note that early studies such as Rayner (1977) and Just and Carpenter (1980) were key studies that demonstrated that readers spend more time looking at words that are used comparatively infrequently in language.[9] While these initial studies did not additionally control for word length, subsequent studies did: these found that, when controlling for differences in word length, the 'word frequency' effect was still strong both in terms of initial fixation on a word, and on gaze duration (Inhoff & Rayner, 1986; Rayner & Duffy, 1986). These findings have subsequently been replicated many times, and further additional nuances have been found:

- High frequency words are skipped more often (Staub & Rayner, 2007);
- Reading time on a low frequency word decreases rapidly if the low-frequency word is subsequently repeated later in the text (Rayner et al., 1995).

There is also a body of research that has demonstrated that time spent reading decreases when word predictability increases. Word predictability in this sense is based on the preceding textual context. Staub and Rayner (2007) noted that word predictability – also known as 'contextual constraint' – is the 'probability that informants will produce the target word as the likely next word in the sentence, given the sentence up to that point' (p. 331). The relationship between predictability and reading time was first noted by Ehrlich and Rayner in 1981 (Staub & Rayner, 2007) and has been supported in subsequent studies.

This research does not directly tell us whether proofreading performance and error detection could be affected by word predictability and word frequency. However, it provides further contextual evidence that lexical factors such as word frequency and word predictability affect reading behaviour, and signals areas for future research.

9 For example, Just and Carpenter (1980) used the Kučera and Francis frequency norms to estimate the normative frequency of words in their experiment.

Discussion

What research has been done on proofreading in relation to text features (in this case, lexical factors)?

This section found and reviewed three studies on proofreading in relation to lexical factors. This constitutes a relatively small number of studies, and the studies reviewed do not include any recent publications or research into lexical factors. This indicates a need for more research into the impact of lexical factors on proofreading.

What is the strength of the existing evidence? To what extent does research into readability and text features have implications for proofreading?

Research has examined lexical factors in relation to proofreading outcomes as well as general reading performance. Although the research (in particular on proofreading) is limited in terms of scope, findings on the impact of lexical factors have been broadly consistent. The research on proofreading outcomes shows that some categories of errors such as homophone foils (e.g. *to-too; see-sea*) are more difficult to detect, while there is some further evidence from both bodies of literature that detection of certain errors is made more difficult by tendencies in the human visual and cognitive systems. For instance, readers respond differently to words of high/low frequency, with common, high-frequency words more likely to be skim-read or skipped by the reader. Reading time also decreases when words are predictable: that is, the reader is more likely in such a scenario to predict a word based on the preceding context rather than read it, which may in turn affect error detection.

These studies allow us to pinpoint and summarise specific lexical features that are more likely to hinder error detection in proofreading. However, the application of these results for improved proofreading practice are discussed only in a very limited way in the literature – or are not treated at all. It is therefore important that further research into commonly undetected errors in proofreading, which stem from specific lexical factors, also test different mitigation strategies that proofreaders might implement.

In this section and elsewhere in this report, the proofreading studies analysed were conducted mostly with undergraduate students as participants, and not professional proofreaders. This may limit the extent to which implications can be drawn for proofreading practice. Two of the proofreading studies reviewed in this section tested participants' reading skills, and found some variation in reading comprehension skills in their samples. While undergraduates can be assumed to have a generally good reading ability – and the variation in ability noted in the studies cited above is therefore on a limited scale only – it is important to acknowledge that the profile of an undergraduate with lower reading comprehension skills might not reflect the profile of a professional proofreader. However, it could be argued that an occasional or informal proofreader will have a comparable reading ability to that of an undergraduate student. For this reason, the recruitment of undergraduates as research participants is not considered a major limitation and provides useful insights for practice.

What possible implications can be drawn to inform proofreading practice?

The methodology adopted to conduct this research included a systematic review of proofreading research and a targeted approach to reviewing readability literature. We found that in both proofreading and readability research, there is little contemporary study on this topic and as such our research did not identify many recently published studies. There is therefore a need for ongoing research into this area in order to draw tangible conclusions.

The preceding discussion leads to a number of tentative considerations for research and proofreading practice.

- There is a need for more empirical research into these effects in the first instance, but also into the effectiveness of possible proofreading strategies to counteract them. For example, a mitigation strategy to investigate could be the use of a ruler, tool or suitable application when reading through text to obscure upcoming text, where this may help to maintain a steady pace of reading and curb the tendency to skim-read high frequency words. This has been found to be a helpful technique in proofreading in L2 and ESL studies, as noted elsewhere in this report. Another beneficial strategy might be to reread sections as errors may not be spotted in a first reading. The proofreader might therefore make a habit of revisiting a text after an initial reading, while organisations might ensure that the time needed to reread is accounted for in overall proofreading time.
- Proofreaders, and organisations employing proofreaders, should be aware that some spelling mistakes are more likely to pass unnoticed than others: for example, homophone errors and errors in high frequency words are less likely to be identified. Equally, there is some evidence that misspelled words that appear to have the correct 'shape' are less likely to be identified.
- A specific avenue for further research is whether technology can assist in screening for circumstances where errors are more likely to occur, and draw the proofreader's attention to these. This could take the form of automatically screening for high frequency words or homophones in a document, and alerting the proofreader to potentially 'high risk' words or sections.

Conclusion and avenues for further research

We examined the literature on proofreading and readability with the aim of understanding the current state of the research, furthering understanding of the impact of text features on proofreading outcomes, and raising awareness of the possible impact of different text features in the design of materials. Throughout this report, key findings from empirical studies on proofreading and readability were reviewed, and to the extent that the strength of the evidence has allowed, possible implications for proofreading practice have been discussed. While one of the intended aims was to generate practical recommendations for everyday proofreading practice, the specific limitations of the empirical research (see section below on 'Limitations') has meant that any recommendations are tentative, and would benefit from further testing to make a stronger case. The ability to make recommendations for proofreading practice is furthermore complicated by the fact that many findings generated in the studies discussed are more easily applied to typesetting and the general preparation of text for proofreading. It is nonetheless intended that this report prompts reflections among both proofreaders and organisations employing them on the existence of materials that are potentially 'high risk' in terms of proofreading outcomes. This may prompt reflection on possible mitigating actions that could be taken when proofreading high risk materials, such as providing proofreaders with additional time for each task or assigning sections of text to multiple proofreaders.

The literature review provided the opportunity to analyse the current state of research in this area. We noted that the literature on some of the topics discussed in this review, such as on optimal text size for reading and proofreading, is coherent and reflects a consensus. On other topics, such as typefaces and specific fonts, the literature is conflicted, where in some instances findings from different experiments were shown to be contradictory. This indicates an important research gap concerning the impact of text features discussed in this report. We reiterate that there is a need for ongoing, contemporary empirical analysis of text features and their impact on proofreading under conditions of modern technology and practice – both of which have changed significantly over time, and will continue to adapt and change over time.

The research method employed in proofreading studies has also seen little evolution over time; yet novel theoretical or empirical approaches could provide new insights into the impact of text features on proofreading outcomes. This may, for example, include qualitative studies that investigate the practices, or perspectives, of professional proofreaders in a more naturalistic environment outside of experimental conditions. While some interactions between (text) features have been examined more closely, such as font and text size interactions, other interactions and knock-on effects have not been the subject of more extensive investigation. For example, the impact of scrolling is discussed in only a limited way in the literature on line features – yet adaptations made to line spacing and line length inevitably increase scrolling when proofreading or reading on-screen, which may disrupt concentration.

A number of the studies examined in this report examined text features and their impact on proofreading performance and readability for text written in languages other than English. In areas where empirical evidence on text features' impact on professional proofreading in English is limited – for example, line length – further research is recommended.

It is also worth exploring if allowing proofreaders some flexibility in personalising the appearance of a document is an effective strategy. In practical terms, this would require training proofreaders on how to customise the appearance of documents to suit their preferences regarding specific text features. We noted some areas where studies reviewed in this report showed limited evidence of text features impacting proofreading or reading performance (e.g. typefaces and fonts), unless considered holistically in relation to other features (e.g. text size). Allowing some customisation in cases where there is not a well-evidenced consistent effect of a text feature on performance would enable organisations employing proofreaders to take subjective preferences into account, arguably without compromising on performance. It is also possible that effects of text features vary according to the individual (i.e. that there is an effect on some individuals but not on others), but that some of these individual differences are 'washed out' in the final results and analysis of empirical studies. Whether customisation is an effective strategy therefore warrants further inquiry while taking these points into consideration.

Finally, while several empirical studies discussed in this report focus on proofreading, very few include suggestions for adaptations to proofreading practice, while even fewer critically assess possible strategies for adapting practice. A so-far unexplored strategy to reduce common errors is the development and use of software to assist in the identification of these types of errors (see section below on 'Software'). In the absence of a more rigorous consideration of mitigation strategies and their possible effectiveness, recommendations for proofreading practice can only be tentative, and further research into such strategies is therefore recommended.

Software

An area for further investigation is whether tailored computer applications or software can reduce common errors in proofreading that relate to text features, for instance by screening for circumstances where errors are more likely to occur and flagging these to the proofreader. This would enable the proofreader to easily be alerted to high frequency words or homophones – such as by highlighting or underlining them – and to do a second check through specific words or parts of the text.

The use of software packages to assist in identifying features of words or sentences has been suggested elsewhere in the readability literature, and may form a useful starting point for investigating the potential uses of software. For example, Beauchamp and Constantinou (2020) used corpus linguistics techniques and software to explore linguistic accessibility in exam questions. In their study, they used AntWordProfiler and Multidimensional Analysis Tagger (MAT) as software packages to identify lexical and syntactic trends, such as low-frequency vocabulary (which they found to be a lexical factor in test items that inhibits reading comprehension). Using software in this way was suggested as a technique to prompt question writers to review material flagged by the software, and seek alternatives when appropriate.

Finally, features of word processing applications, such as the Immersive Reader function in Microsoft 365, may be useful in enabling proofreaders to manipulate text features and to perform other functions such as obscuring parts of the text while proofreading. Examining the effectiveness of allowing proofreaders more scope in document customisation would be a useful area of further inquiry.

Limitations

While some of the research on text features in relation to reading performance has examined the effect on *proofreaders* in terms of error detection and the proofreading experience, proofreading studies make up only a small proportion of the overall literature analysed in this review. Although we argue that the readability literature has relevance for understanding the impact of text features on proofreading, the relatively small number of empirical studies on proofreading outcomes must still be acknowledged as a limitation. In addition, some areas of the proofreading literature discussed in this report (e.g. lexical factors) have seen little development in recent decades, and, in the absence of significant recent contributions to these debates, there is limited scope to make robust claims.

In many of the empirical studies on proofreading and readability, participants recruited in experiments were primarily undergraduate students rather than professional proofreaders. On the one hand, undergraduate students with a good command of the language they are proofreading reflect the skill level of many who engage in proofreading in desk-based roles on an ad-hoc or informal basis. However, the lack of experienced or professional proofreaders among the research participants in these studies acts as a minor limitation on the extent that direct or tangible implications for professional proofreading can be drawn.

The methodological features of the studies reviewed often revealed a focus on specific factors and their interactions, making generalisations difficult. Of the proofreading papers treated in this report, some have a specific focus, while others describe experiments that take place under specific conditions. For example, some studies examined interaction effects between text features and other specifically defined factors (e.g. background noise) that may not have wide applicability to typical conditions of proofreading. In other instances, experiments were conducted in a way that the experimental conditions were artificial and not entirely reflective of real-world or everyday proofreading tasks, such as proofreading very short passages. In addition, this report clearly specifies where experiments have been conducted in a language other than English. In some cases, studies examined the impact of formatting features that are highly specific to a language or alphabet (e.g. text direction in Chinese proofreading), and these features were excluded from the analysis of this report. Findings from proofreading studies in different languages may have limited generalisability to languages beyond than the ones they have been tested on, and recommendations on text features such as optimal line length and line spacing (print, or on-screen) may vary between languages and alphabets.

References

Arps, R. B., Erdmann, R. L., Neal, A. S., & Schlaepfer, C. E. (1969). Character legibility versus resolution in image processing of printed matter. *IEEE Transactions on Man-Machine Systems*, *10*(3), 66-71. https://doi.org/10.1109/TMMS.1969.299885

Bean, W., & Bouffler, C. (1987). *Spell by Writing*. Primary English Teaching Association.

Beauchamp, D., & Constantinou, F. (2020). Using Corpus Linguistics Tools to Identify Instances of Low Linguistic Accessibility in Tests. *Research Matters*, *29*, 10-16.

Beier, S. (2009). *Typeface Legibility: Towards defining familiarity*. Royal College of Art.

Bernard, M. L., Chaparro, B. S., Mills, M. M., & Halcomb, C. G. (2003). Comparing the effects of text size and format on the readibility of computer-displayed Times New Roman and Arial text. *International Journal of Human-Computer Studies*, *59*(6), 823-835. https://doi.org/10.1016/S1071-5819(03)00121-6

Burt, C., Cooper, W., & Martin, J. (1955). A psychological study of typography. *British Journal of Statistical Psychology*, *8*(1), 29-56. https://doi.org/10.1111/j.2044-8317.1955.tb00160.x

Chan, A. H., & Ng, A. W. (2012). Effects of display factors on Chinese proofreading performance and preferences. *Ergonomics*, *55*(11), 1316-1330. https://doi.org/10.1080/00140139.2012.714472

Chan, A. H., Tsang, S. N., & Ng, A. W. (2014). Effects of line length, line spacing, and line number on proofreading performance and scrolling of Chinese text. *Human factors*, *56*(3), 521-534. https://doi.org/10.1177/0018720813499368

Cogie, J., Strain, K., & Lorinskas, S. (1999). Avoiding the proofreading trap: The value of the error correction process. *The Writing Center Journal*, *19*(2), 7-32.

Cohen, G. (1980). Reading and searching for spelling errors. In U. Frith (Ed.), *Cognitive processes in spelling* (pp. 135-155). Academic Press.

Creed, A., Dennis, I., & Newstead, S. (1987). Proof-reading on VDUs. *Behaviour & Information Technology*, *6*(1), 3-13.

Crisp, V., Johnson, M., & Novaković, N. (2012). The effects of features of examination questions on the performance of students with dyslexia. *British educational research journal*, *38*(5), 813-839.

Daneman, M., & Stainton, M. (1991). Phonological recoding in silent reading. *Journal of experimental psychology: Learning, Memory, and Cognition*, *17*(4), 618. https://doi.org/10.1037/0278-7393.17.4.618

Dyson, M. C. (2004). How physical text layout affects reading from screen. *Behaviour & Information Technology*, *23*(6), 377-393. https://doi.org/10.1080/01449290410001715714

Dyson, M. C., & 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*(4), 585-612. https://doi.org/10.1006/ijhc.2001.0458

Dyson, M. C., & Kipping, G. J. (1998). The effects of line length and method of movement on patterns of reading from screen. *Visible language*, *32*(2), 150.

Ganayim, D., & Ibrahim, R. (2013). How do typographical factors affect reading text and comprehension performance in Arabic? *Human factors*, *55*(2), 323-332.

Gould, J. D., Alfaro, L., Barnes, V., Finn, R., Grischkowsky, N., & Minuto, A. (1987). Reading is slower from CRT displays than from paper: Attempts to isolate a single-variable explanation. *Human factors*, *29*(3), 269-299.

Grabinger, R. S., & Osman-Jouchoux, R. (1996). Designing screens for learning. In H. v. Oostendorp & S. d. Mul (Eds.), *Cognitive Aspects of Electronic Text Processing* (pp. 181-212). Greenwood Publishing Group.

Halin, N., Marsh, J. E., Haga, A., Holmgren, M., & Sörqvist, P. (2014). Effects of speech on proofreading: can task-engagement manipulations shield against distraction? *Journal of Experimental Psychology: Applied*, *20*(1), 69. https://doi.org/10.1037/xap0000002

Healy, A. F., Volbrecht, V. J., & Nye, T. R. (1983). The effects of perceptual condition on proofreading for misspellings. *Memory & cognition*, *11*(5), 528-538. https://doi.org/10.3758/BF03196990

Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & psychophysics*, *40*(6), 431-439. https://doi.org/10.3758/BF03208203

Jared, D., Levy, B. A., & Rayner, K. (1999). The role of phonology in the activation of word meanings during reading: evidence from proofreading and eye movements. *Journal of Experimental Psychology: General*, *128*(3), 219.

Jones, D. M., Miles, C., & Page, J. (1990). Disruption of proofreading by irrelevant speech: Effects of attention, arousal or memory? *Applied Cognitive Psychology*, *4*(2), 89-108. https://doi.org/10.1002/acp.2350040203

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: from eye fixations to comprehension. *Psychological review*, *87*(4), 329. https://doi.org/10.1037/0033-295X.87.4.329

Köpper, M., Mayr, S., & Buchner, A. (2016). Reading from computer screen versus reading from paper: does it still make a difference? *Ergonomics*, *59*(5), 615-632.

Levy, B. A., Di Persio, R., & Hollingshead, A. (1992). Fluent rereading: Repetition, automaticity, and discrepancy. *Journal of experimental psychology: Learning, Memory, and Cognition*, *18*(5), 957. https://doi.org/10.1037/0278-7393.18.5.957

Ling, J., & Van Schaik, P. (2006). The influence of font type and line length on visual search and information retrieval in web pages. *International Journal of Human-Computer Studies*, *64*(5), 395-404.

Ling, J., & van Schaik, P. (2007). The influence of line spacing and text alignment on visual search of web pages. *Displays*, *28*(2), 60-67. https://doi.org/10.1016/j.displa.2007.04.003

Lonsdale, M. (2014). Typographic features of text. Outcomes from research and practice. *Visible language*, *48*(3), 29-67.

Luckiesh, M., & Moss, F. (1940). Boldness as a factor in type-design and typography. *Journal of applied psychology*, *24*(2), 170. https://doi.org/10.1037/h0061254

Macinska, S., & Pastorino, C. (2020, October 2020). *Embedded features for accessible on-screen assessment: A review of the evidence on effectiveness* Education Assessment Research Seminar, Cambridge Assessment,

McMullin, J., Varnhagen, C., Heng, P., & Apedoe, X. (2002). Effects of surrounding information and line length on text comprehension from the web. *Canadian Journal of Learning and Technology/La revue canadienne de l'apprentissage et de la technologie*, *28*(1). https://doi.org/10.21432/T27607

Monk, A. F., & Hulme, C. (1983). Errors in proofreading: Evidence for the use of word shape in word recognition. *Memory & cognition*, *11*(1), 16-23. https://doi.org/10.3758/BF03197657

Nanavati, A. A., & Bias, R. G. (2005). Optimal line length in reading-A literature review. *Visible language*, *39*(2), 120.

Nihei, Y., Terashima, M., Suzuki, I., & Morikawa, S. (2002). Why are four eyes better than two? Effects of collaboration on the detection of errors in proofreading. *Japanese Psychological Research*, *44*(3), 173-179. https://doi.org/10.1111/1468-5884.00020

Piepenbrock, C., Mayr, S., & Buchner, A. (2014). Positive display polarity is particularly advantageous for small character sizes: implications for display design. *Human factors*, *56*(5), 942-951. https://doi.org/10.1177/0018720813515509

Porte, G. (2001). Losing sight of errors: the effects of typographical conditions on error salience in L2 proof-reading. *System*, *29*(1), 137-148. https://doi.org/10.1016/S0346-251X(00)00050-6

Rayner, K. (1977). Visual attention in reading: Eye movements reflect cognitive processes. *Memory & cognition*, *5*(4), 443-448. https://doi.org/10.3758/BF03197383

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & cognition*, *14*(3), 191-201. https://doi.org/10.3758/BF03197692

Rayner, K., Raney, G. E., & Pollatsek, A. (1995). Eye movements and discourse processing. In R. F. Lorch & E. J. O'Brien (Eds.), *Sources of Coherence in Reading* (pp. 9–36). Erlbaum.

Shaikh, A. D., & Chaparro, B. S. (2005a, April 2005). *The effect of line length and passage type on reading speed, comprehension, and user satisfaction* 1st Annual Symposium on Graduate Research and Scholarly Projects, Wichita, KS.

Shaikh, A. D., & Chaparro, B. S. (2005b). The Effects Of Line Length On Reading Performance Of Online News Articles. Proceedings of the Human Factors and Ergonomics Society Annual Meeting,

Sheedy, J. E., Subbaram, M. V., Zimmerman, A. B., & Hayes, J. R. (2005). Text legibility and the letter superiority effect. *Human factors*, *47*(4), 797-815. https://doi.org/10.1518/001872005775570998

Spencer, H. (1969). *The visible word* (2nd ed.). Royal College of Art.

Staub, A., & Rayner, K. (2007). Eye movements and on-line comprehension processes. In M. G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (Vol. 327, pp. 342). Oxford University Press.

Szafranek, K. (2009). *Font Smoothing Explained*. Retrieved 22 February 2009 from https://szafranek.net/blog/2009/02/22/font-smoothing-explained/

Tinker, M. A. (1963). *Legibility of print*. Iowa State University Press.

Tinker, M. A., & Paterson, D. G. (1939). Influence of type form on eye movements. *Journal of Experimental Psychology*, *25*(5), 528. https://doi.org/10.1037/h0063038

Tullis, T. S., Boynton, J. L., & Hersh, H. (1995). *Readability of fonts in the windows environment*. https://doi.org/10.1145/223355.223463

Venetjoki, N., Kaarlela-Tuomaala, A., Keskinen, E., & Hongisto, V. (2006). The effect of speech and speech intelligibility on task performance. *Ergonomics*, *49*(11), 1068-1091. https://doi.org/10.1080/00140130600679142

Vitello, S., & Mouthaan, M. (2022). *What impacts success in proofreading? A literature review of proofreading on screen vs on paper.* Cambridge University Press & Assessment Research Report.

Wilkins, A., Smith, K., & Penacchio, O. (2020). The influence of typography on algorithms that predict the speed and comfort of reading. *Vision*, *4*(1), 18. https://doi.org/10.3390/vision4010018