CAMBRIDGE
UNIVERSITY PRESS & ASSESSMENT

What impacts success in proofreading? A literature review of proofreading on screen vs on paper

Research Report

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:

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

Acknowledgements

I wish to thank David Beauchamp for conducting the preliminary literature searches and review of the journal articles. I also want to thank Melissa Mouthaan for engaging in initial discussions with me about the empirical findings and themes emerging from the review.

Contents

Executive Summary

Context

- There is no doubt that there are many advantages of proofreading on screen, especially considering the environmental impacts of printing and how central computers have become to our work and daily lives. However, there continues to be a screen versus paper debate among proofreading communities. Various proofreading guidance documents, blogs and online commentaries still discourage proofreading on screen. To what extent is this backed by research evidence?
- To answer that question, we conducted a systematic literature review of the empirical studies published in peer-reviewed journals that have directly compared proofreading performance (accuracy or speed) when it was conducted on screen versus on paper.
- This report contains a comprehensive overview and analysis of all the studies reviewed, including a set of tables that highlight key methodological details and key findings. This can act as a reference document to ensure that discussions and decisions around proofreading procedures are research-informed.

What empirical studies have been conducted?

- In total we found 13 journal articles, which reported on 26 experiments that met our review criteria.
- Almost all the experiments measured both proofreading accuracy and speed. In most cases, speed was measured as reading speed rather than the speed of detecting errors. Some experiments also included measures of psychological or physiological effects (e.g., reading comfort).
- As a whole this is a relatively large evidence base, but it is primarily made up of relatively old studies – published in the 1980s and 1990s. Only four studies were published after 2000 (seven experiments in total).
- Across the experiments, there was a large amount of variation in certain features of their methodologies, especially the screen types used, the movability of the paper, topics of the materials being proofread, and the duration of the proofreading tasks. Therefore, we cannot rule out that some differences between studies may be due to the specific decisions regarding the experimental set-up. In contrast, there was little variation in other aspects, such as the types of materials being proofread (all used written text), the types of participants (mostly university students) and the proofreading task (identifying errors verbally). Together, this may limit the generalisation or ecological validity of their findings to contexts not studies (e.g., proofreading of text in images or speech transcripts).

What does the evidence show?

- The results were mixed for both accuracy and speed. Various studies have reported statistically significant differences between proofreading on screen compared to on paper, with effects primarily in the direction of screen disadvantages. But, overall, more experiments have failed to find significant effects. In addition, for speed (but not accuracy), the finding of disadvantages seemed to have reduced over time; all were

reported in 1980s and 1990s with none afterwards. Accuracy effects have been inconsistent in all decades.
- There was some evidence of apparent mismatches between participants' performance and their perceptions of proofreading under screen and paper conditions. This may explain why some proofreaders hold the belief that their performance will be worse on paper (if they have a worse experience when proofreading on screen).
- The research literature appeared limited with regard to the explanations or mechanisms underlying the screen disadvantages (or lack of). Researchers have been unable to convincingly identify the conditions under which proofreading on screen is worse (or better) than on paper.
- Several researchers have put forward hypotheses for screen effects on proofreading performance. Visual-based hypotheses have been dominant since the 1980s for explaining screen disadvantages on speed. The idea is that material displayed on screen appears visually different than on paper. Whilst this has received converging support from various experiments, it cannot fully explain all the empirical findings, especially not effects on accuracy. Non-visual-based explanations have also been proposed), but none have not been thoroughly investigated yet.
- It is possible that no adequate explanations for screen versus paper effects have been found yet because proofreading has not been sufficiently considered from the perspective of a complex human activity. Every proofreading activity is conducted by a proofreader under a dynamic and interacting set of psychological, social, physical, and environmental conditions. Yet, previous research has mainly focused on investigating proofreading from a cognitive processing perspective, and on isolating specific factors rather than exploring their interactions.
- The methods used in the studies may also have limited their capability to identify the conditions under which screen is likely to be worse (or better) than paper for proofreading. We noted methodological shortcomings, for example, in how studies tested for moderating effects on performance, in how they engaged with proofreading as a complex activity, and in the ecological validity of their proofreading set-up.

Final reflections

- Overall, the empirical research we reviewed provided no conclusive evidence that proofreading on screen will, in general, lead to worse performance than proofreading on paper, especially when conducted under screen conditions typically used nowadays (e.g., using computers that display high quality visual images). Therefore, this does not support claims that proofreading should be conducted on paper as the default.
- Together, the evidence base points to the need to conduct more research. What is needed is to have an adequate theory (theories) that can predict with high certainty what conditions will lead to effective proofreading on screen. On the one hand, research could devote more attention to understanding the specific conditions that can improve or worsen proofreading on screen relative to paper. On the other hand, research could focus on understanding how to improve proofreading on screen in its own right.

Recommendations

- Despite various questions remaining unresolved, we put forward three practical recommendations from this research base:

- If poorer performance is found during on-screen proofreading, do not assume that it is due to screen use. Investigate this thoroughly, as there are many factors that may affect proofreading performance, some of which may interact, and the screen may not always be a primary factor.
- Ensure the screens and materials being displayed have high visual quality. The evidence suggests that currently used computer screens seem to be good enough. However, image quality can be impaired for various reasons other than the screen itself. For example, zooming in or out a PDF can make the text appear clearer or fuzzier. Images can have poorer or lower resolution, which may reduce the text quality.
- Explore how to use screen tools (e.g., spelling and grammar checkers, zooming, text masking) to enhance the proofreading experience, rather than trying to equate the experience of proofreading on screen to proofreading on paper.

Introduction

Proofreading and proofreading errors

Proofreading is a necessary part of getting a document ready for sharing with others. Often we think of proofreading as the process of checking a final version of a document for any remaining errors, especially of spelling, punctuation and grammar. But it can have a wider remit than this and can occur at earlier stages in the production of a document (Mouthaan & Vitello, 2022).

> A professional proofreader should find and correct almost all of these in a text: spelling errors…serious, unarguable errors of punctuation, especially where they allow ambiguity or obscure the meaning; inconsistently spelled or hyphenated names; bad word breaks that make reading the text difficult; incorrect text headings and page headers/footers …incorrect page numbers and cross-references; missing text; repeated text; wrongly placed or incorrect captions and annotations. (Chartered Institute of Editing and Proofreading, 2022).

In Cambridge University Press & Assessment, proofreading is integral to ensuring the quality of a diverse range of products and outputs, including assessments, syllabus documents, qualification certificates, textbooks, academic books, teaching and learning resources, research reports, conference presentations, and many more. It is important that we continually review how effective our proofreading processes are because proofing errors can impact quality, validity and reputation in a variety of ways. For example, spelling, punctuation and grammar (SPaG) errors in question papers can turn an answerable question into an unanswerable one.

If we understand what factors affect proofreading, then we can implement improvements that should increase the likelihood that errors are detected by proofreaders. However, proofing errors, like other types of errors such as in content and marking, can occur for a complex set of reasons (Suto & Ireland, 2021; Vitello & Rushton, 2021). In a recent report on error in assessment materials (Vitello & Rushton, 2021), we discussed one model for categorising different causes of error, which applies both to why errors are introduced as well as why they are not detected during checks. This model is based on the aviation industry's SHELLO model and was modified based on our research into errors in assessment materials and assessment practices. Our modified version, called the SHELLOP model, distinguishes between seven different types of factors that contribute to the risk of error, as follows:

1. **Software**: Computer programmes and any documents that are used.
2. **Hardware**: Tools or equipment that are used.
3. **Environment**: Features of the internal or external environment.
4. **Liveware** (people): Characteristics of individual people involved in the process.
5. **Liveware-liveware** (people-people): Interpersonal and social aspects.
6. **Organisation**: Organisational structure, processes and culture.
7. **Product**: Characteristics of the product (or service) being worked on

This current report is the second of two research reports that have aimed to evaluate the research evidence into physical aspects of the equipment or materials that proofreaders interact with that may affect proofreading success. Under the SHELLOP model of error in assessment materials (Vitello & Rushton, 2021), these aspects can be viewed as relating to "software", "hardware" and "product" categories of risk. For the first report (Mouthaan & Vitello, 2022), we conducted a literature review on the visual features of text to understand the extent to which different texts may be more or less 'risky' with regard to error detection. For this second report, we conducted a literature review on the debate between proofreading on screen compared to on paper. Psychological, social and environmental factors are also important to understand proofreading performance, but in this set of reports they are only discussed with regard to their relationship and interaction with the physical aspects of the proofreading process.

The ongoing debate: screen versus paper

One recurrent debate, which has become particularly timely again, is the potential difference in effectiveness between proofreading on paper compared to proofreading on a computer screen. Traditionally, documents have been proofread on paper rather than on screen even when this has required the extra effort of printing out a copy of the document. It is important to revisit this debate now because of various recent and anticipated changes to ways of working, life and education. These have created both push factors, which make proofreading on paper less appealing than before, as well as pull factors, which make proofreading on screen more attractive. Below are some examples:

- *Environmental commitments.* Proofreading on screen can be more environmentally beneficial than proofreading on paper, as it removes the carbon footprint of paper production and printing. However, the precise environmental benefit of not needing to use paper always needs to be weighed against the carbon emissions of computer use. On-screen proofreading also has a (digital) carbon footprint, which depends on a variety of factors (e.g., screen time, type of computer equipment, data sharing infrastructure).
- *Ways of working*. Since 2020, the COVID-19 pandemic has required many of us to work from home rather than the office, which has affected access to printed versions of documents.
- *Digital educational products*. Many educational products, including some assessments, are being developed for a digital platform only, which negates one of the common reasons for proofreading on paper – to proofread the document in the same way the intended user would.
- *Increased use of computers*. In work and everyday life, many of us are doing many more tasks on the computer, including reading and editing documents.
- *Technology upgrades*. Computer hardware and software have improved greatly over the past several decades. Modern computer screens, for example, have much higher resolution than before and have reduced some undesirable features (e.g., flickering).
- *Ways of communicating*. Proofreading on screen can facilitate how proofreading comments are recorded, tracked and communicated to authors, especially in ways that may minimise the risk that proofreaders' comments are misunderstood or (unintentionally) overlooked by the author.

Despite the potential advantages of on-screen proofreading, there continues to be various examples of guidance documents, blogs or other online commentaries that discourage proofreading on-screen and encourage proofreaders to print out documents (e.g., see College of Media and Publishing, 2022; Cruickshank, 2021; Plain English Campaign, 2022).

Some are based on anecdotal accounts from professional proofreaders while others are claims that seem to be presented as fact and are not accompanied with citations to the research evidence on which the claims were based. Other organisations do not appear to recommend one mode over the other and provide supporting guidance for both (e.g., Chartered Institute of Editing and Proofreading, 2022).

This current report

In this current report we focused on the empirical research on proofreading. Our key question was: to what extent does the empirical evidence show that proofreading is more (or less) successful when it is conducted on paper compared to on screen?

We conducted a literature review of the research on proofreading that compared paper and on-screen performance. We reviewed the research findings specifically from the perspective of proofreading, which was not always the aim of those empirical studies. Many studies that tested proofreading performance used proofreading tasks as a way to understand impacts on reading (not proofreading) or impacts of working with computers in general. Our distinct aim was to understand whether the evidence suggests that proofreading specifically is impacted differently by the display mode (screen vs paper). This required us to carry out a detailed inspection of the following: the methodology of the studies (e.g., the tasks, the participants, the measures of performance), how these methodological features compared to our work contexts, and the level of consistency between the empirical findings.

In order to answer the research question directly, we primarily focused on studies that (i) directly compared screen and paper proofreading (i.e., they had to have compared both conditions in the study), and (ii) tested accuracy or speed as measures of proofreading success. A few studies tested speed without testing accuracy, which arguably is a less useful as a measure of success, given that something done quickly could be done less accurately. However, we reasoned it was important to include these studies as they provided another type of evidence that we could use to evaluate the consistency of the effects and the potential conditions under which proofreading may be more, or less, successful.

Methodology

Literature search

We searched the academic research literature for empirical studies that directly compared proofreading on screen with proofreading on paper. The literature search was conducted as part of wider literature review on factors affecting proofreading success, and therefore followed the same approach as described in Mouthaan and Vitello (2022) which looked specifically at effects of text features. In summary, we carried out an initial thorough search of the literature in in late 2020 using Google Scholar as a literature database, and then subsequent searches on ERIC and Scopus. We did not apply any date filters in order to capture all research studies that have been conducted. We used a variety of key terms during our search, such as 'proofreading', 'proofreading errors; 'proofreading accuracy', 'proofreading speed', 'proofreading performance', and containing at least one of the following key terms 'screen', 'on screen', 'paper, 'print', 'computer' 'display'. We re-ran these searches in 2022 to identify any studies that had been published after the initial search, but no other studies were found.

We inspected the abstracts and methods sections of all these papers for the following inclusion criteria. The studies had to:

i. be published in a peer-reviewed journal.
ii. describe at least one experiment using a proofreading task.
iii. include a direct comparison between at least one screen and one paper condition.
iv. measure at least one measure of proofreading success – accuracy or speed.

In total 13 journal papers were found that met all the above criteria. We could not obtain access to one of the papers (Oliver, 1993), and therefore we had to exclude it from the review. The remaining 12 papers were fully reviewed.

Interpreting findings within the context of the publication date

During the literature review, it became apparent that the studies needed to be reviewed within the context of their publication date. This is because significant changes have happened in computer technology over time. We needed to take these changes into account when drawing conclusions about present-day effects of screen use – this point has been emphasised by many researchers of the articles we reviewed (e.g., Köpper et al., 2016).

Examples of changes to computer screen technology include: increased display resolution, background luminance, and refresh rates; flicker-free displays; functionality for changing display polarity (screen background and text colour); functionality for presenting the same fonts as can be printed; introduction of anti-aliasing[1] and sub-pixel rendering of text; screen size; and screen movability. In addition, we need to remember that these improvements have been occurring alongside societal, attitudinal and behaviour changes around computers, such as increased access, usage and expertise.

1 Anti-aliasing techniques smooth the jagged edges of images (including letters and numbers). Aliased technology present "characters and lines that appear to contain "staircasing" or "the jaggies."… caused by an undersampling of the signal that would be required to produce sharp, continuous characters." (Gould, Alfaro, Finn, et al., 1987, p.499).

Köpper et al. (2016) drew a specific distinction between on-screen reading studies conducted prior to 2000 and those after 2000s due to the type of computer technology used. They noted that most studies in the 1980s and 1990s used now-obsolete older computer technology (cathode ray tube (CRT) displays), which suffered from low display resolution and low luminance as well as flickering, especially the oldest models. Also, early computer operating systems, especially in the 1980s, were limited in terms of the visual features they could present on screen. For example, some computers presented text in negative polarity only (light text on dark), could only display particular character fonts and had jagged pixilation around the text due to a lack of anti-aliasing techniques. Although there has continued to be improvements in screen and computer technology after 2000, we think that the year 2000 is still a useful reference point for distinguishing older from newer screen technology. Therefore, in this report we have often discussed studies and findings in relation to whether it was published before or after 2000.

Studies from both time periods are useful, although in different ways. Post-2000 studies help us evaluate the likelihood that screen effects may occur nowadays, as screen technology used in these studies is far more similar to what is used by current proofreaders than in studies before 2000. In contrast, those earlier studies help us to understand the extent to which computer technology may play a role in proofreading success, and to be aware of screen conditions that may particularly problematic so that we ensure those features are eliminated where possible.

Findings

This section is organised around two key questions that guided our literature review and analysis. Together they help us understand the current state of research evidence on the effects of display mode on proofreading success.

1. What empirical studies have been conducted?
2. What have these empirical studies found?

What empirical studies have been conducted?

Overview and summary

We reviewed 12 research articles that described experiments that directly compared at least one screen and one paper condition for proofreading performance. 26 experiments were reported in total. The earliest article was published in 1983 (Wright & Lickorish, 1983) and the last was published in 2017 (Hargis et al., 2017). There were various gaps of around 10 years where no research meeting our search criteria was published (between 1987 and 1998; between 1998 and 2008 and between 2008 and 2015).

Out of the 26 experiments we reviewed, 25 measured proofreading accuracy and 24 measured speed. In almost all cases, speed was measured as reading speed (e.g., number of words read during a time limit) rather than the speed of detecting errors. Some of the studies also examined other variables such as perceptions of proofreaders' own accuracy or self-reported physical or psychological symptoms (e.g., fatigue, reading comfort, eyestrain). In addition, the studies varied with regard to the computer technology they used, the set-up of the paper condition and the experimental procedure. However, the studies provided different amounts of detail about key features including about the paper and screen conditions, the participants, the experimental materials, the tasks and procedure. This meant that we could not obtain a perfect understanding of the empirical basis of all the findings, although, in most cases, there was sufficient information to evaluate and compare the studies.

In “Appendix A – Methodological details of the studies reviewed”, we provide three tables with key information about every study we reviewed. Table A - 1 contains details of the characteristics of the participants, proofreading materials and tasks used in the studies. Table A - 2 focuses on several key features of the screens and paper used in the studies (screen type, display size and resolution and positioning). Table A - 3 provides details on two features of text presentation - polarity and text font types – as these features have been highlighted as having key effects on reading and proofreading (Mouthaan & Vitello, 2022). Below we provide a summary of the methodological similarities and differences from those tables.

Participants, proofreading materials, and tasks (Table A - 1)

Participants

The studies tested between 6 to 79 participants in each display condition, although most studies used samples of around 20 participants. Those that used the smallest samples (6-10 participants) were often testing specific experimental manipulations as part of a muti-experiment study (Gould, Alfaro, Barnes, et al., 1987). Köpper et al. (2016) used much larger samples than any of the other studies (more than 60 participants). In addition, we noted that across the 30 years of experiments there seems to have been a change in the characteristics of participants. All the studies after 2000 tested relatively young participants (mean age in 20s or 30s), and most were university students. This means that it is unlikely that they would have had any professional proofreading training. While university students participated in some earlier studies too, studies before 2000 used participants with a wider range of background, including participants from a Psychology research panel (Wright & Lickorish, 1983), research professionals, and clerical staff (Gould, Alfaro, Barnes, et al., 1987; Gould, Alfaro, Finn, et al., 1987; Gould & Grischkowsky, 1984). In addition, in some of the earliest studies, the researchers explicitly noted that their participants had (or were likely to have) little experience of using computer screens (e.g., Gould & Grischkowsky, 1984; Wright & Lickorish, 1983).

Proofreading materials

Most studies asked participants to proofread only one or two texts per display condition. Köpper et al. (2016) stood out for using a substantially larger number of texts – 7 to 21 texts per display condition. In addition, the studies used different types of text, although details were often missing about the genre, source, length or layout of the texts. Most texts were non-fiction texts (e.g., newspaper articles, popular science texts, college and university texts) with some experiments using short stories. Different topics were used (e.g., smoking, the music industry, animals). The texts also differed in length; some studies used texts of a single page whilst others used lengthy multi-page texts (e.g., 10 pages), and the number of words ranged from around 200 words to more than over 1500 words. There was slightly more consistency in the types of errors planted in the texts (to be detected by the participants). Almost all studies included spelling errors either as the only type of error or in combination with other errors such as errors of grammar, punctuation, formatting or repeated words. Shibata et al. (2015) and Gujar et al. (1998) did not use spelling errors at all, choosing semantic errors or rhyming substitution errors (e.g., “cake” replaced with “fake”), respectively. Across the studies, these planted errors varied in how often they appeared in the texts; errors occurred much less frequently in some texts than in others (e.g., one error approximately every 170 words vs one every 15 words).

Proofreading task

Across the decades there also seem to have been changes to the nature of the proofreading tasks that have been used, especially with regard to the time participants spent proofreading as well as how participants were asked to record errors they detected. Earlier studies tended to give participants longer to proofread (10 to 50 minutes per text) than seen in more recent studies (2 to 4 minutes). To some extent, this is because articles tended to be longer in earlier studies, although participants were, in most cases, not required to complete the text in the allocated time. In addition, some of the earlier studies did not give participants any time limit; for example, Creed et al. (1987) simply asked participants to read the text once while Wright and Lickorish (1983) asked participants to self-pace without any restriction.

Regarding the specific task, in four of the earliest experiments participants were asked to indicate errors differently when proofreading on paper than on screen. When proofreading on paper participants marked directly on the paper but when proofreading on screen they either marked the errors on screen (Creed et al., 1987; Wright & Lickorish, 1983), on a separate paper (Wright & Lickorish, 1983) or pointed to the error on the screen (Gould & Grischkowsky, 1984). However, most experiments since then have removed this task difference by asking participants to indicate errors verbally for both display conditions. Making the responses equivalent between conditions is important from an experimental perspective, but it does not reflect what typically happens during proofreading (errors are noted on the text). It was rare to find studies that asked participants to correct the errors they detected or to be permitted to make annotations as they proofread.

Features of the screen and paper conditions (Table A - 1, Table A - 2, Table A - 3)

There were missing details for many studies, mostly for studies conducted before 2000.

The computer screens

There was a lot of variation in the computer technology that was used across the experiments we reviewed. Table A - 2 shows which screen types (e.g., CRT, LCD), sizes, resolution, and screen positions were used whilst Table A - 3 shows the different polarity (text vs background colour) and font types used to display the text. Unsurprisingly several features showed a pattern of technological improvement over time. For example, the availability of certain screen types changed after around 2000 which affected the characteristics of the displays (Köpper et al., 2016). All the pre-2000 studies we reviewed used CRTs and all those after 2000 used LCDs. Two studies used tablets (Köpper et al., 2016; Shibata et al., 2015). Screen resolution tended to be higher in the more recent studies than in earlier studies.

In addition, the computer technology in very early studies placed restrictions on the polarity and font types that could be used to display the text. Very early studies in the 1980s used computers, for example, that displayed green or red text on dark backgrounds, or dark text on “light grey” or “white-ish” backgrounds. When polarity became controllable, all studies displayed black text on white background rather than the reverse (i.e., positive rather than negative polarity). Font types similarly were restricted in early computers used in the experiments but, unlike polarity, no consistent font emerged across the later studies. Instead, they have used a variety of font types (e.g., Helvetica, Times Roman) and font sizes. The size of screens also varied across the studies, although unlike other features, there did not appear to be a consistent trend towards increasing or decreasing size across time.

One other feature of the screens that stood out in our review was the position or movability of screens. Most studies pre-2000 did not explicitly mention how they positioned the screen, but we can assume that they either could not be adjusted (as was the case for early computer screens) or were adjusted to a default orientation. However, studies from around the end of the 1990s studies started to describe the positions explicitly. Gujar et al. (1998) stated they fixed their screens 15° from vertical while Köpper et al. (2016) fixed their non-tablet screens 75° backwards and their tablet screen 15°. Shibata et al. (2015) also used a tablet but did not fix its position, allowing participants to move it freely.

Screen vs paper comparability

When comparing proofreading performance between screen and paper conditions, it is important to understand how the screen characteristics compared to the set-up of the paper condition. For most studies, we found it difficult to determine exactly how equivalent the conditions were because they simply did not provide as much detail for the paper condition as for the screen conditions. We could not find any information at all about paper size, resolution or position for several studies.

Despite these reporting issues, there seemed to be two approaches that have been taken to address screen vs paper comparability across the studies we reviewed. On the one hand, there were experiments that attempted to remove as many differences as possible between the screen and paper conditions. Sometimes taking this option was not under the control of the researchers. This was the case for the earliest studies, where limitations in computer technology prevented researchers from ensuring the screen and paper had the same set-up. However, since improvements in technology, most studies (if not all) have removed many of the now-controllable differences in text presentation (e.g., polarity and text font). On the other hand, there were experiments that were set-up intentionally to maintain certain differences between the conditions that reflected differences in how they would be used in practice (i.e., outside the experimental laboratory setting). The main features of the set-up that have continued to differ between recent studies are those related to the positioning and movability of the screen and paper when proofreading. Across the years, some studies allowed participants to freely move the paper when proofreading whilst in other studies the paper was restricted in a similar set-up to the screen condition (e.g., paper was placed in a wooden frame, on a copy holder, desk stand or in casing).

What have these empirical studies found?

Overview and summary

Table 1 and Table 2 show a summary of the findings from all the empirical studies we reviewed. Overall, the evidence was mixed. Across the last three decades, many experiments have reported significant screen disadvantages on accuracy or speed, but a similar number of experiments have found no significant differences relative to proofreading on paper. In addition, some studies have reported both of these types of findings within the same experiment. In some cases, this was because the studies tested different screen and paper conditions and found that the screen disadvantage only appeared when certain screen and paper conditions were compared, but not for all comparisons (e.g., Gould, Alfaro, Finn, et al., 1987). In others, the mixed effects appeared because the studies tested for interactions between the display manipulation (screen vs paper) and other task manipulations (e.g., the text topic), and the screen disadvantages only appeared in certain combinations of those manipulations (e.g., Wharton-Michael, 2008).

Across all the studies, only two studies have reported screen advantages. In one of them (Köpper et al., 2016), the screen advantage was found in a specific experimental set-up where a tablet was used and it only facilitated speed not accuracy. In the other study (Gould, Alfaro, Finn, et al., 1987), screen advantages were found for both accuracy and speed, but only when higher-quality screen conditions were compared to poorer-quality screen or poor-quality paper conditions.

Within this mix of findings, there were some screen effects that appeared more consistent than others. For example, screen effects have been found more often for speed than for accuracy and a cluster of studies have reported participants having worse perceptions of the proofreading experience when proofreading was conducted on screen. However, these more consistent effects on speed and perceptions were more typically seen in studies published before 2000 and seem to have largely disappeared in later studies. Indeed, no study after 2000 has reported screen disadvantages on speed, although it must be noted that far fewer studies have been conducted after 2000 than before. Screen effects on accuracy, in contrast, have been mixed across the three decades of research.

Table 1. Overview of empirical findings[2] from all studies that directly compared proofreading on screen vs on paper.

Red and blue highlight significant paper and screen advantages, respectively; purple shows where both significant screen and paper advantages were found within the same study; darker shades highlight more consistency. See Appendices B and C for more details of the comparisons and results (e.g., means).

Study	Accuracy		Speed	
	Better (mean)	**Were differences statistically significant (sig.)?**	**Faster (mean)**	**Were differences statistically significant (sig.)?**
Hargis (2017)	Paper	Yes	Speed not tested	
Köpper (2016) E1	Screen	No	Screen	No
Köpper (2016) E2	Screen	No	Paper	No
Köpper (2016) E3	Screen	No	Mixed	No
Köpper (2016) E4	Screen	No	Screen	Yes
Shibata (2015) E1	Paper	Yes	Paper	No
Wharton-Michael (2008)	Paper	Mixed (sig. and non-sig paper advantages)	Speed not tested	
Gujar (1998) E1	?	No	?	No
Ziefle (1998) E1	Paper	Yes	Paper	Yes
Gould (1987b[3]) E1	Mixed	No	Paper	Mixed (sig. and non-sig paper advantages)
Gould (1987b) E2	Paper	No	Paper	No
Gould (1987b) E3	Paper	No	Mixed	No
Gould (1987b) E4	Mixed	No	Mixed	No
Gould (1987b) E5	Paper	No	Paper	Mixed (sig and non-sig paper advantages)
Gould (1987b) E6	Mixed	Mixed (sig. screen advantages, non-sig. screen advantages, non-sig paper advantages)	Mixed	Mixed (sig. screen advantages, sig. paper advantages, non-sig. paper advantages)
Gould (1987a[4]) E2	Mixed	No	Paper	Yes
Gould (1987a) E3	Paper	No	Paper	Yes
Gould (1987a) E6	Accuracy not tested		Paper	Yes
Gould (1987a) E7	Paper	Yes	Paper	Yes
Gould (1987a) E8	Screen	No	Paper	Yes
Gould (1987a) E9	Screen	No	Paper	No
Wilkinson (1987)	Paper	Yes	Paper	Yes
Creed (1987) E1	Paper	Mixed (sig. and non-sig. paper advantages)	Mixed	No
Creed (1987) E2	Mixed	Mixed (sig. paper advantages, non-sig screen advantage)	Paper	Yes
Gould (1984)	Paper	No	Paper	Yes
Wright (1983)	Paper	Mixed (sig and non-sig. paper advantages)	Paper	Yes

2 Studies have multiple results because they compared multiple screen and paper conditions or tested for interactions, and found different effects. See Appendix B for more details.
3 Gould, J. D., Alfaro, L., Finn, R., Haupt, B., & Minuto, A. (1987). Reading from CRT displays can be as fast as reading from paper. Human factors, 29(5), 497-517.
4 Gould, J. D., Alfaro, L., Barnes, V., Finn, R., Grischkowsky, N., & Minuto, A. Ibid.Reading is slower from CRT displays than from paper: Attempts to isolate a single-variable explanation. (3), 269-299.

Table 2. Brief summary of other effects of display mode (not accuracy or speed) that have been investigated by the studies reviewed.

Study	Effects of display mode (other than accuracy or speed)
Hargis (2017)	No differences in perceptions of proofreaders' own accuracy.
Köpper (2016) E1	No differences for most self-reported physical or psychological symptoms. Screen increased fatigue and eyestrain.
Köpper (2016) E2 and E3	No differences for most self-reported physical or psychological symptoms. Screen increased eyestrain and reduced reading comfort.
Köpper (2016) E4	No differences for all self-reported physical or psychological symptoms[5].
Shibata (2015) E1	Differences in self-reports such as feeling less able to concentrate or performing the task quickly and more fatigue on screen. No difference in self-reported ability to detect errors. Differences in physical interactions with screen and paper.
Gujar (1998) E1	Differences in self-report preferences and experiences - paper preferred overall.
Ziefle (1998) E1	Differences in self-report preferences and experiences - paper preferred overall.
Gould (1987b[6]) E1-E6	Differences in self-report preferences and experiences - paper preferred overall.
Gould (1987a[7]) E6	Eye movements – mixed findings.
Gould (1987a) E8	Differences in self-report preferences and experiences.
Gould (1984)	No differences on measures of comfort, vision or body movements.

Accuracy – why are the screen effects so mixed?

The literature review revealed a mixed picture in terms of accuracy. First, looking at the results of the significance tests, screens either appeared to significantly worsen accuracy or to have no significant effect on accuracy (relative to proofreading on paper), although there have been more non-significant than significant differences reported. Second, looking at the mean differences additionally showed that the non-significant results were inconsistent with regard to whether the paper or screen condition was numerically more accurate. The paper condition was not always more accurate; several studies have found higher accuracy in screen conditions, although only one of these differences reached statistical significance.

Therefore, this raised the question as to why these findings are so mixed. To what extent can these inconsistencies help us understand the conditions under which using the screen for proofreading might affect accuracy?

This set of findings provides us with two important points for the screen vs paper debate on accuracy. First, the overall lack of consistency in screen effects on accuracy indicates that simply using a screen cannot be the primary driver of poor accuracy; if this were the case, then most (if not, all) studies should have reported screen disadvantages. Second, the fact that mixed results were found within all decades of research means that accuracy effects are also unlikely simply to be due to screen features that were more inferior in earlier years of computers (e.g., poorer screen resolution) or screen unfamiliarity (i.e., lower levels of computer experience in the population). However, this does not mean that these factors do not have any impact on accuracy – they may interact with each other or with other features of the proofreading setup to increase, minimise, or even eliminate screen effects completely.

5 Screen differences were for the agitation measure, but there were pre-experiment differences between the screen and paper participants (between-subject design)

6 Gould, J. D., Alfaro, L., Finn, R., Haupt, B., & Minuto, A. (1987). Reading from CRT displays can be as fast as reading from paper. Human factors, 29(5), 497-517.

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

To understand which factors may moderate screen effects on accuracy, we first turn to studies that have experimentally manipulated features of the proofreading set-up. 23 of the 26 experiments we reviewed also investigated moderating factors. Table 3 shows that six experiments found some evidence of significant moderators of the screen versus paper difference on accuracy. Interestingly, most of these factors were not related to characteristics of the display, concerning instead the topic of the proofreading material, duration of the proofreading task, type of proofing errors, and annotation method. Across all the factors, two have only been investigated once (topic of the proofreading text and annotation method) and, therefore, we cannot determine how reliable their effects are. The other factors have been explored, to some extent, in more than one experiment (duration of the proofreading task, type of proofing errors, visual characteristics of the display, visual quality of the display), but none have shown consistent results.

Visual characteristics of displays have been investigated the most across the studies we reviewed. However, only two studies have found evidence that visual factors may moderate the difference between screen and paper accuracy, and neither one was able to identify a specific characteristic contributing to effect. Creed et al. (1987) compared proofreading on screen with proofreading from a printed paper as well as proofreading from a photograph of the screen display so that the material looked more similar to how it was presented on screen. There was a significant screen disadvantage relative to the printed paper condition only, but the photograph condition did not differ significantly from either the screen or print condition, which suggests that the screen disadvantage effect was not strong. Gould, Alfaro, Finn, et al. (1987) found some evidence suggesting that the visual quality of the display more generally could moderate the screen versus paper difference. In Experiment 6, they compared three screen types with different combinations of display characteristics (e.g., resolution, anti-aliasing of characters, character style, polarity, spacing of lines) and two paper conditions that differed in visual quality (good-quality print and very poor-quality print where the appearance of the characters were distorted). This experiment showed that visual quality can create a screen advantage when the quality of the paper is poor. They found no difference between the different screens they tested in terms of accuracy, but proofreading on any of the screens was either the same as proofreading on the good quality paper or it was better than proofreading on poor quality print. None of those effects on accuracy, however, were replicated across the other experiments that Gould, Alfaro, Finn, et al. (1987) conducted. Various experiments have tried to disentangle the different visual characteristics that could be moderating screen effects, but with little success. These experiments have investigated display luminance, display resolution, display regeneration rate, display contrast, display polarity, display orientation and display aspect ratio. The impact of physical characteristics of the display were also explored in one experiment, by encasing the proofreading paper in different cases, but this also failed to affect proofreading accuracy relative to the screen condition.

Given the lack of consistent evidence for any specific moderating factor, it is useful to turn our attention to the set of studies that have found significant screen effects on accuracy. This may reveal characteristics of the experimental set-up that might explain why these specific studies found differences between screen and paper proofreading. In the studies before 2000 we noticed that there was one particular research team (Gould and his colleagues at IBM) that consistently reported no significant effects on accuracy. As this team did consistently report speed effects, it suggests that their method was not insensitive to all screen effects. However, it is not clear why their findings differed from most of the others in the same time period. We cannot rule out the possibility that some findings of worse accuracy in the 1980s, and even 1990s, were simply due to low computer familiarity.

After 2000, three studies have reported significant effects of display on accuracy (Hargis et al., 2017; Shibata et al., 2015; Wharton-Michael, 2008). Shibata et al. (2015) inspected how participants interacted with the screen and paper during proofreading to determine if that could explain their finding of worse accuracy under the screen condition. Unlike most other studies, they allowed participants to annotate and interact with the screen like they would with paper, by using a tablet rather than desktop screen. They found that participants pointed to text and slid their finger/pen across the material significantly more when proofreading from paper than from the screen, and that this significantly correlated with error detection rates. While this is a novel finding, some caution is needed when interpreting it. First, according to Shibata et al. (2015), some participants had difficulty in interacting with (touching) the tablet, which seemed to be due to unfamiliarity with the tablet set up. This may have exacerbated the screen disadvantage. Second, text touching was not experimentally manipulated, and therefore it is not possible to determine whether text touching caused the lower error rates or was a by-product of another unknown causal factor.

We also had reservations about some methodological aspects of the other two studies that found accuracy disadvantages after 2000. Wharton-Michael's (2008) study contained a confounding variable between its screen and paper conditions: participants marked proofing changes to the materials differently when proofreading was conducted on screen (using track changes on Microsoft Word) compared to on paper (write on the paper). This task difference may have contributed to the screen effects on accuracy, although it may have been the more natural way of proofreading in both instances. Finally, it was impossible to evaluate Hargis et al. (2017)'s methodology to the same level of depth because they provided few details about the screen condition. This condition was only briefly mentioned in the discussion, as it had been included as part of a follow-up experiment to check the conclusion of the main study about age effects on metacognition.

Table 3. Experiments that investigated potential moderating factors of screen vs paper differences [8] [9]

Study	Significant paper v screen		Factors moderating screen vs paper difference[10] (*)		
	Acc.	Speed	Significant effect on accuracy	Significant effect on speed	No significant effects for speed or accuracy
Köpper (2016) E2	=	=	-	-	Proofreading duration (*indirect vs E1)
Köpper (2016) E3	=	=	-	-	Display luminance (*screen)
Köpper (2016) E4	=	✓	-	Display's ergonomic characteristics (*screen, indirect vs E1-3)	-
Wharton-Michael (2008)	x =	n/a	Text topic	n/a	Error placement
Gujar (1998) E1	=	=	-	-	Display resolution (*screen); Display's visual characteristics (*paper); Display's physical characteristics (*paper)
Ziefle (1998) E1	x	x	-	-	Display resolution (*screen)
Gould (1987b) E1	=	x =	-	-	Font type
Gould (1987b) E2	=	=	-	Display resolution (*screen, indirect vs E1)	Page length
Gould (1987b) E3	=	=	-	-	Page length; Display regeneration rate (*screen)
Gould (1987b) E4	=	=	-	Polarity (screen)	Display contrast (*screen)
Gould (1987b) E5	=	x =	-	Anti-aliasing image (*screen)	-
Gould (1987b) E6	✓=	x ✓=	Display's visual quality (*screen, *paper)	Display's visual quality (*screen, *paper); Anti-aliasing image (*screen)	-
Gould (1987a) E2	=	x	-	-	Orientation (*paper)
Gould (1987a) E3	=	x	-	-	Aspect ratio (*paper)
Gould (1987a) E7	x	x	-	-	Display's visual characteristics (*paper)
Gould (1987a) E8	=	x	-	-	Display's visual characteristics (*screen)
Gould (1987a) E9	=	=	-	-	Polarity (*paper); Aspect ratio (*paper)
Wilkinson (1987)	x	x	Proofreading duration	-	Practice
Creed (1987) E1	x =	=	Display's visual characteristics (*paper)	-	Error type
Creed (1987) E2	x =	x	Error type	-	Text format: columns
Gould (1984)	=	x	-	-	Proofreading duration
Wright (1983)	x =	x	Annotation method	Annotation method x practice	-

8 x means a significant screen disadvantage; ✓means a significant screen advantage; = is means no significant difference between screen and paper found
9 Some studies have multiple results because they either compared multiple screen and paper conditions or tested for display x task interactions, and found different effects.
10 We have highlighted when manipulations were implemented either only in the screen or paper condition, and when the effect of the manipulation was evaluated against a condition from a different experiment (i.e., an indirect comparison). Both limit the conclusions we can draw about these effects.

Proofreading speed – why have screen disadvantages disappeared over time?

For proofreading speed, there was a relatively consistent pattern of screen versus paper differences across the studies we reviewed. First, all the significant effects on speed that have been found (except one) were screen disadvantages, relative to paper. Second, all these screen disadvantages have been reported in studies published before 2000. None have been reported after 2000. Third, the screen disadvantages reported by these early studies were found in almost every study conducted in that time period. There was one exception that stood out in the review. Köpper et al. (2016) Experiment 4 was the only one to report a significant screen effect among the studies published after 2000, and, in this case, they found a screen advantage on speed, relative to paper. This screen advantage has not been replicated in any other studies and, therefore, we cannot determine how reliable it may be.

Taken together, this set of findings shows that the screen disadvantages on speed seem to have disappeared over time. The raises the question of why and whether the empirical evidence can explain this apparent shift. As we did for accuracy, we turn our attention to the experiments (within our review) that investigated potential factors that could moderate screen differences on speed (Table 3).

From Table 3 we can see that there is little consistent evidence of the specific factors that moderate the screen versus paper difference on speed. 23 of the experiments we reviewed investigated this, but only six found evidence of any moderating effects on speed. Most of these effects seemed to be due to visual characteristics of the display (display resolution, anti-aliasing, display visual quality), one concerned the display's ergonomic characteristics (incline position of the screen) and the other was not related to display characteristics at all (it was an interaction between the annotation method and task practice). However, none of these factors individually has strong empirical support when looking across the entire set of studies.

Some of the moderating effects have only been tested in one experiment, which makes it impossible to determine how robust the findings are (display ergonomic characteristic; annotation method x practice effect, visual quality). One effect has been directly replicated across two experiments, which is the finding that screens with anti-aliased images are proofread as quickly as paper whereas screens without anti-aliasing are proofread more slowly than paper (Gould, Alfaro, Finn, et al., 1987 Experiments 5 and 6). But this finding is also somewhat tenuous, as the replication was conducted as part of the same multi-experiment study, rather than by an independent research team. Display resolution is the other factor that has been tested in multiple studies, but it has only shown evidence of a moderating effect in one study, and this was only indirect evidence, as it was shown via a comparison with another condition in a different experiment (Gould, Alfaro, Finn, et al., 1987 Experiment 2).

In addition, many other factors have been explored across the studies we reviewed but without producing any significant effects on moderating the screen disadvantage on speed. These factors have concerned visual features of the display (luminance, regeneration rate, polarity, aspect ratio), physical features of the display (movability of the paper), text presentation (font type, page length, text column format), and task features (type of proofing errors, proofreading task duration).

What can explain this apparent conflict between the overall pattern of disappearing effects over time and the lack of strong evidence of moderating factors?

As with any null findings, it is possible there was low statistical power to detect small moderations on speed or less consistent effects within groups of participants. There is, however, another explanation, which was suggested by Gould and colleagues in the late 1980s after running two multi-experiment studies to try to disentangle the screen speed disadvantage (Gould, Alfaro, Barnes, et al., 1987; Gould, Alfaro, Finn, et al., 1987). Gould, Alfaro, Finn, et al. (1987) concluded that their evidence, as a set, suggested "the explanation centers on the image quality of the characters." (p. 497).

In the first of these two studies, Gould, Alfaro, Barnes, et al. (1987) conducted 10 experiments that aimed to "isolate a single-variable explanation" (p. 269) of the screen speed disadvantage. In summary, they investigated what they called "task variables", "display variables", and "personal variables" (p.271). Six of these experiments included a direct comparison between proofreading speed on screen and on paper (see Table 3) and compared different characteristics of the screen and/or paper set-up to determine if they impacted the screen vs paper difference. They examined display orientation, display aspect ratio, display polarity, and the display's visual characteristics more generally (e.g., via comparing different screen types). The other experiments produced other evidence to help explain the screen effect. Experiment 1 and 5 compared different paper conditions (Experiment 1 tested for paper visual quality effects while Experiment 5 tested for visual angle and font type effects). Experiment 10 compared a paper condition with a pseudo-screen condition. Experiment 4 did not measure effects on speed, aiming, instead, to find out whether task differences between screen and paper proofreading could explain the speed difference; they specifically examined if there were differences in reading distances when proofreading on screen compared to on paper. They also looked at effects of experience of screen reading and age in some experiments. These experiments were not able to identify any single variable that accounted for the speed difference being found between screen and paper proofreading and led to the following conclusion:

> We tentatively believe that the difference may be due to a combination of several variables. … Most of the evidence, including that from later experiments (Gould, Alfaro, Finn, Haupt, and Minuto, in press), suggests that the image quality of the characters, rather than task or user variables, is most likely responsible for the reading-speed difference. (p.297).

This second study (referred to in the quote above) took a different approach to investigating the screen speed disadvantage. Gould, Alfaro, Barnes, et al. (1987) explained they had "turned from looking for an explanation of the reading-speed difference to searching for conditions in which people can read as fast from CRT displays as from paper." (p. 298). In the second study, six experiments were run, all with a direct screen and paper comparison (Gould, Alfaro, Finn, et al., 1987). Already in their first experiment, Gould, Alfaro, Finn, et al. (1987) found evidence of a reduced screen speed disadvantage than that seen in previous experiments they had run. They suggested that this was because the screens had used improved technology which enabled them to present the proofreading materials in a more similar way to paper than previous screen technology had been able to do. "This experiment compares, for the first time, proofreading performance on paper and on CRT when the material proofread has the same font, polarity, size, color (almost), and layout on the two media." (p. 499). The next set of experiments tested different screen and paper conditions to determine whether they could determine the specific conditions under which proofreading can be as fast on screen as on paper. They examined a range of display factors (resolution, regeneration rate, polarity, contrast, anti-aliasing of images) and general visual quality of the displayed materials by comparing different screen types with different combinations of characteristics, or different paper conditions that had been manipulated to be worse in image quality than other paper. After all the experiments, Gould, Alfaro, Finn, et al. (1987) made the “tentative conclusion” that “it appears that display polarity (dark characters on a light, whitish background), improved display resolution, and anti-aliasing itself each contributes to the elimination of the paper/CRT display reading rate difference. Associated with these three variables was a difference in fonts, but we are unable to make an argument independent of the three aforementioned variables to the effect that font itself contributed to this reduction.” (p. 516). They went on to argue that “the behavioral explanation for the improved reading from CRT displays is probably visual rather than cognitive or emotional. This is because we have demonstrated that the variables contributing to improved proofreading from CRT displays are visual in nature.” (p.516).

Visual quality explanations for speed disadvantages cannot be the complete explanation, however. A later study by Ziefle (1998) showed a screen disadvantage despite using similar screens to Gould, Alfaro, Finn, et al. (1987).

Other effects – what is the evidence?

Several experiments we reviewed investigated whether other aspects of the proofreading experience (beyond accuracy or speed) were differentially affected by proofreading on screen compared to on paper (see Table 2).

Many studies before 2000 reported differences in participants’ self-reported preferences and experiences of proofreading on screen and on paper. Paper proofreading was consistently preferred overall, over screen proofreading. However, there were still some early studies that reported no differences on measures of comfort, vision, body movements or eye movements (Gould, Alfaro, Finn, et al., 1987; Gould & Grischkowsky, 1984). Similarly, the more recent study by Shibata et al. (2015) also found that participants’ reported having negative experiences of proofreading such as feeling less able to concentrate and being more fatigued when proofreading screen. The rest of the most recent studies have not reported many differences between participants’ experiences of screen and paper. Köpper et al. (2016) asked participants to report on many physical and psychological symptoms. The only significant findings were that screen increased symptoms of fatigue (found in Experiment 1), increased eyestrain (in Experiment 1, 2 and 3) and reduced reading comfort (Experiment 2 and 3). In Experiment 4, all symptomatic differences seemed to have been eliminated, which Köpper et al. (2016) attributed to the use of a screen tablet that was inclined to the same position as paper. However, without a direct comparison with other screen orientations, this conclusion is tenuous.

In addition, these other effects often did not show a consistent relationship with the performance effects. For example, in all of Gould, Alfaro, Finn, et al. (1987)’s experiments, participants preferred to proofread on paper, but the findings showed no consistent differences in their participants’ performance either on accuracy or speed in most cases. Similarly, Köpper et al. (2016)’s participants reported increased physical discomfort (e.g., eyestrain) in three of their experiments but none showed any significant differences in accuracy or speed. Hargis et al. (2017) shows a different example of this misalignment.

They asked participants to predict their own accuracy. Despite no perceived differences in accuracy, they found that participants proofread significantly worse (missing more errors) on screen than on paper.

Discussion

The aim of this literature review was to understand what empirical evidence is available to inform the debate between proofreading on screen versus on paper. This is important because, despite the nowadays routine use of computers for many work and non-work activities, there still appears to be a strong belief, within some organisations and proofreading communities, that proofreading is more successful when it is conducted on paper than on screen (e.g., College of Media and Publishing, 2022; Cruickshank, 2021; Plain English Campaign, 2022). To what extent is this supported by evidence?

What empirical studies have been conducted?

Our review found 26 experiments that have directly compared proofreading success under screen and paper conditions, where success was measured in terms of accuracy or speed. Some of these experiments also included measures of psychological or physiological effects (e.g., participants' reading comfort and display preferences). Although, as a whole, this is a relatively large evidence base, it is primarily made up of old studies – published before 2000. We found only four studies post 2000; these were published between 2008 and 2017, and reported on seven experiments in total. There was a large amount of variation in certain features of the methodologies employed, especially with regard to the screen types used, the movability of the paper, topics of the materials being proofread, and the duration of the proofreading tasks. Therefore, we cannot rule out that some differences between studies may be due to the specific decisions regarding the experimental set. In contrast, there was little variation in other aspects of the methodology, including the types of materials being proofreading (all used written text), the types of participants (mostly university students) and the proofreading output (identifying errors verbally only). Together, this may limit the generalisation or ecological validity of their findings to contexts not studies (e.g., proofreading of text in images or speech transcripts).

What has the evidence shown?

Looking across the entire set of studies, the results were mixed for both accuracy and speed. Various studies have reported significant differences between proofreading performance on screen compared to on paper, and these effects have been primarily in the direction of screen disadvantages (rather than advantages). But, overall, more experiments have failed to find significant effects than those that reported significant differences. In addition, the evidence has shown apparent mismatches between participants' performance and their perceptions of proofreading under screen and paper conditions. Negative perceptions of proofreading on screen were often not accompanied by poorer performance relative to proofreading on paper. This may explain why some proofreaders hold the belief that their performance will be worse on paper (because they have a worse experience when proofreading on screen).

Although mixed findings were evident for both accuracy and speed, they were inconsistent in different ways. When it came to accuracy, there was no evidence of any consistent screen differences across studies. Within each decade of research, some studies found screen disadvantages and others did not. In contrast, for speed, the level of consistency of findings seemed to change over time. All the screen speed disadvantages were reported in 1980s and 1990s, and most within the earliest set of studies. After 2000, only one experiment found significant screen effects, and this was in the opposite direction – screen advantage. This fact that recent studies have repeatedly failed to find screen effects (at least on speed) provides some support for the conclusion that has been drawn by various researchers that proofreading on screen can be as successful a proofreading on paper, which has been attributed to technological advances and societal changes around computer usage (Gould, Alfaro, Finn, et al., 1987; Köpper et al., 2016).

Having an evidence base of mixed findings can be theoretically and practically useful when it helps identify the specific conditions that lead to different effects. Unfortunately, the research we reviewed was relatively limited in that respect. For both accuracy and speed, a small number of experiments showed some evidence of factors significantly moderating the screen disadvantage, but no single factor had strong empirical support. There also appeared, once again, to be differences in the patterns between accuracy and speed. The evidence seemed to show a tendency for non-display features to moderate accuracy effects but visual display factors to moderate speed effects more often, which suggests that the same mechanism may not be involved for both. However, this is simply a hypothesis at this stage, as this apparent difference has not been investigated directly and the evidence is not large enough to draw stronger conclusions.

Theoretical gaps within the research literature

We could not find any theoretical explanations within the literature we reviewed that appeared able to adequately explain the entire set of empirical findings of screen versus paper effects on accuracy or speed. However, several researchers have suggested possible explanations, which are considered in more detail below.

The most common explanation, discussed in many of the studies we reviewed, is the notion that proofreading differences may be caused by differences between the visual characteristics of screen and paper conditions (Creed et al., 1987; Gould, Alfaro, Barnes, et al., 1987; Gould, Alfaro, Finn, et al., 1987; Wilkinson & Robinshaw, 1987; Wright & Lickorish, 1983; Ziefle, 1998). This visual-based hypothesis was particularly prominent in studies before 2000, presumably because older computer technology limited how material could be displayed on screen, which resulted in obviously different visual characteristics to paper. The strongest support for this hypothesis has come from two sets of experiments conducted by Gould and colleagues (Gould, Alfaro, Barnes, et al., 1987; Gould, Alfaro, Finn, et al., 1987).

Visual-based explanations are still being used by researchers when discussing the debate between screen and paper. Köpper et al. (2016) used this to explain why none of their four experiments showed significant screen differences relative to paper on accuracy or speed.

> Most likely, this reflects the improved characteristics of modern TFT-LCD screens. For instance, screen resolution in Experiment 1 (128 ppi) was considerably higher as compared with the screen resolution in earlier studies…Furthermore, the TFT-LCD technology provides a flicker-free image, the texts were displayed in positive polarity, and the letters were anti-aliased using sub-pixel rendering. All of these factors may have contributed to an improved display quality on the computer screen and therefore have increased proofreading speed and performance in comparison to earlier research. (p.621).

However, when looking at the entire set of empirical findings, visual-based explanations seem lacking in broader support. First, these explanations are only partial, in that no researcher has yet specified the mechanism by which visual features may worsen proofreading performance on screen. At the time Gould, Alfaro, Finn, et al. (1987, p.516) expressed caution about this, labelling their visual-based conclusion as “tentative” and acknowledging that they were “uncertain about what visual mechanisms might be responsible for this overall effect, aside from a general realization that the characters simply are now more easily discriminated from one another” (Gould, Alfaro, Finn, et al., 1987, p.516). Attempts to isolate the effects of specific visual factors on the screen disadvantages have been largely unsuccessful (Gould, Alfaro, Barnes, et al., 1987; Ziefle, 1998) or not consistently replicated (e.g., the font type effects reported by Creed et al., 1987). Second, visual-based explanations cannot seem to account for the fact that screen disadvantages on speed and accuracy have been found in later studies that seemed to use similar (Ziefle, 1998) or better (Shibata et al., 2015) quality screens than Gould, Alfaro, Finn, et al. (1987). Third, several studies have found various non-display factors moderating screen disadvantages, especially on accuracy.

Non-visual-based explanations have also been proposed by some researchers but, again, without usually being accompanied by specific mechanisms of how proofreading performance may be differently affected by screen versus paper displays. One example of specific mechanisms being discussed can be found in Shibata et al. (2015). When trying to explain why the lower amount of “text touching” (pointing and sliding fingers/pen over text) on screen relative to paper may correlate with better error detection, they suggested the following possible mechanisms:

> If people are unable to point to text, they might forget to devote attention to a certain part, or it might become difficult to compare distantly positioned parts of text. If they are unable to slide a finger or a pen on text, people might skip words or sentenced during reading. It is possible that such inconvenience lowers reading performance. (p.568-569).

Despite these reasonable suggestions, the evidence we reviewed, however, does not provide support for text-touching as a general explanation of the screen versus paper difference. Various studies we reviewed constrained participants’ ability to interact with the paper or screen displays when proofreading and yet this does not seem to correlate in any obvious way to the findings. Indeed, Köpper et al. (2016) compared proofreading on a tablet screen and paper, where text touching was restricted on both, and they found a significant advantage of using the tablet on speed.

It is possible that no adequate explanations for screen versus paper effects have been found yet because proofreading has not been sufficiently considered from the perspective of a complex human activity. The SHELLOP model of complex human systems (Suto & Ireland, 2021; Vitello & Rushton, 2021) explains that complex activities result from the interactions between characteristics of the person doing the activity (e.g., psychological, emotional, physical), features of the software and hardware used to perform task, characteristics of the context in which the activity is conducted (e.g., environment, social, organisational processes and culture), and task features that place demands on how the activity needs to be conducted (e.g., nature of the proofreading material). In various proofreading studies we reviewed, especially early ones, researchers have acknowledged the complexity that might be involved, and that different factors may combine in non-obvious ways to affect the screen disadvantage relative to paper (Creed et al., 1987; Gould, Alfaro, Barnes, et al., 1987).

Creed et al. (1987) suggested that the display font used in their screen conditions could explain the screen disadvantages, but they also cautioned about the “simplicity of this explanation”. They argued that “it would be unwise to ignore the other factors which affect performance, such as the user's attitude to the VDU [Visual Display Unit], the user's motivation, and the user's working environment” (p. 12). However, this has not yet led to research exploring this complexity.

Methodological limitations of the existing research

Our review also revealed that the methods used in the experiments may have, in various ways, limited their capability to determine the reasons behind screen versus paper differences on performance. We discuss three aspects as examples: moderating effects, theoretical complexity, and ecological validity.

The first aspect concerns the methods used to test for factors that may moderate screen disadvantages. A strong experimental design to test for moderating effects will fully cross the factors being tested; this means, that all combinations of the factors will be compared. Most experiments, especially after 2000 or for certain factors (visual, physical or ergonomic display characteristics) have not done this, using indirect approach instead. For example, some studies compared conditions in one experiment to those in previous experiment or they manipulated the factors within either the screen or paper conditions, rather than both. Of course, sometimes direct comparisons may have been difficult, or even impossible to conduct for practical limitations in modifying the screen or paper set-up.

The second methodological aspect concerns how to explore complex human activities. As discussed earlier, it is likely that proofreading is a complex task, where various factors of different types interact in determining proofreading success. To date, studies have examined a small number of potential factors that could affect performance, having focused largely on visual characteristics of the displays. This complexity has also been suggested to explain why it has been difficult to find evidence of moderating factors:

> Proofreading is probably a complex task, because participants might use various reading strategies involving different combinations of cognitive and visual factors (e.g., reading for comprehension or scanning mainly for unfamiliar letter clusters and word shapes). It is likely that different reading strategies caused the high variance among participants, possibly obscuring effects of CRT resolution on reading performance. Thus, effects of display resolution should be investigated using a task that requires observers to scan information in a more homogeneous way. (Ziefle, 1998, p.560).

There is also an argument to be made about introducing more ecological validity into experiments on proofreading to substantially enhance the screen versus paper debate. For example, in the studies we reviewed participants were often university students rather than professional proofreaders. The tasks were not authentic proofreading tasks as they did not typically require participants to correct the errors, annotate or make comments on the materials, which is usually required in professional contexts. The screen condition did not make use of proofreading support tools that exist such as mark-ups and other technology that could facilitate the task (e.g., zoom, highlighting, spell checkers, etc). It is possible that this kind of research is being conducted outside of academia, within proofreading communities or organisations (we limited our review to studies published in peer-reviewed journals). Related to this is the finding that only certain types of materials were proofread in these studies (paragraphs of written text). There are many other types of materials that have text needing to be proofread (e.g., graphs, data tables) and other formats in which the materials are presented (e.g., text on websites or in brochures).

Final reflections and practical recommendations

Overall, the empirical research we reviewed provided no conclusive evidence that proofreading on screen will, in general, lead to worse performance than proofreading on paper, especially when conducted under screen conditions typically used nowadays (e.g., using computers that display high quality visual images). Therefore, this does not support claims that proofreading should be conducted on paper as the default.

However, the evidence also suggests that there will be some cases when proofreading is likely to be more successful on paper than on screen, and some when the reverse may be true (screen proofreading being better than paper proofreading). As yet, there seems to be no adequate theory that appears able to predict with high certainty what these specific conditions are. Screen technology, especially displays' visual quality, has dominated explanations of screen differences, but even this is unable to explain why some studies that have used similar screens have found effects on accuracy or speed while others have not.

Together, the evidence base points to the need to conduct more research. There are different ways future research could proceed. On the one hand, research could devote more attention to understanding the specific conditions that can improve or worsen proofreading on screen relative to paper. For this, we argue that proofreading should be viewed as a complex human activity to ensure that explanations adequately consider the potential interactions of a comprehensive set of factors (e.g., psychological, social, environment) in affecting performance. On the other hand, research could focus on understanding how to improve proofreading on screen in its own right. By solely comparing proofreading on screen with proofreading on paper, this may limit our capability to explore the unique advantages that computers may give us for conducting proofreading tasks (Mouthaan & Vitello, 2022).

Despite various questions remaining unresolved about proofreading on screen versus on paper, we can make some practical recommendations from this evidence base.

- If poorer performance is found during on-screen proofreading, do not assume that it is due to screen use. Investigate this thoroughly, as there are many factors that may affect proofreading performance, some of which may interact, and the screen may not always be a primary factor.

- Ensure the screens and materials being displayed have high visual quality. The evidence suggests that currently used computer screens seem to be good enough. However, image quality can be impaired for various reasons other than the screen itself. For example, zooming in or out of a PDF can make the text appear clearer or fuzzier. Images can have poorer or lower resolution, which may reduce the quality of any text on the image.

- Explore how to use screen tools (e.g., spelling and grammar checkers, zooming, text masking) to enhance the proofreading experience, rather than trying to equate the experience of proofreading on screen to proofreading on paper.

References

Chartered Institute of Editing and Proofreading. (2022). *What is proofreading?* Chartered Institute of Editing and Proofreading. Retrieved 5th May from https://www.ciep.uk/about/faqs/what-is-proofreading/

College of Media and Publishing. (2022). *Proofreading on-screen tips*. College of Media and Publishing. Retrieved 5th May from https://collegeofmediaandpublishing.co.uk/blog/proofreading-screen-tips/

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

Cruickshank, L. (2021). *Five rules for proofreading your own work: Proofing pinciples to help make sure your writing shines*. Scottish Book Trust. Retrieved 5th May from https://www.scottishbooktrust.com/articles/5-rules-for-proofreading-your-own-work

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.

Gould, J. D., Alfaro, L., Finn, R., Haupt, B., & Minuto, A. (1987). Reading from CRT displays can be as fast as reading from paper. *Human factors*, *29*(5), 497-517.

Gould, J. D., & Grischkowsky, N. (1984). Doing the same work with hard copy and with cathode-ray tube (CRT) computer terminals. *Human factors*, *26*(3), 323-337.

Gujar, A. U., Harrison, B. L., & Fishkin, K. P. (1998). A comparative evaluation of display technologies for reading. Proceedings of the Human Factors and Ergonomics Society Annual Meeting,

Hargis, M. B., Yue, C. L., Kerr, T., Ikeda, K., Murayama, K., & Castel, A. D. (2017). Metacognition and proofreading: the roles of aging, motivation, and interest. *Aging, Neuropsychology, and Cognition*, *24*(2), 216-226.

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.

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

Oliver, R. (1993). Proof-Reading on Paper and Screens: The Influence of Practice and Experience on Performance. *Journal of Computer-Based Instruction*, *20*(4), 188-124.

Plain English Campaign. (2022). *Twelve tips for proofreading*. Plain English Campaign. Retrieved 5th May from http://www.plainenglish.co.uk/proofreading.html

Shibata, H., Takano, K., & Tano, S. i. (2015). Text touching effects in active reading: The impact of the use of a touch-based tablet device. IFIP Conference on Human-Computer Interaction,

Suto, I., & Ireland, J. (2021). Principles for Minimizing Errors in Examination Papers and Other Educational Assessment Instruments. *International Journal of Assessment Tools in Education*, *8*(2), 310-325.

Vitello, S., & Rushton, N. (2021, September 2021). *From flying a plane to creating exam papers: how the SHELLO model can help us minimise errors in assessment materials.* Education Assessment Research Seminar, Cambridge University Press & Assessment (online).

Wharton-Michael, P. (2008). Print vs. computer screen: Effects of medium on proofreading accuracy. *Journalism & Mass Communication Educator*, *63*(1), 28-41.

Wilkinson, R., & Robinshaw, H. M. (1987). Proof-reading: VDU and paper text compared for speed, accuracy and fatigue. *Behaviour & Information Technology*, *6*(2), 125-133.

Wright, P., & Lickorish, A. (1983). Proof-reading texts on screen and paper. *Behaviour & Information Technology*, *2*(3), 227-235.

Ziefle, M. (1998). Effects of display resolution on visual performance. *Human factors*, *40*(4), 554-568.

Appendix A – Methodological details of the studies reviewed

Table A - 1. Summary of participants, proofreading materials and tasks in studies that compared screen (S) and paper (P) conditions

Experiment (E)	Participants		Materials proofread			Proofreading task	
	N[11]	Characteristics	N[11]	Characteristics	Errors per text	Time per text	Task
Hargis (2017)	31	Mean age 21	2 texts	Non-fiction texts from Wikipedia or a GRE exam[12]; varied topics; mean 236 words.	16 misspellings and grammar errors	3 min	P – Errors circled S –
Köpper (2016) E1	66-70	Most university students; mean age 23	7 texts	Short stories; mean 870 words, 2 columns.	16 misspellings and 14 grammar errors	3 min	S&P - Errors indicated verbally
Köpper (2016) E2	79	Most university students; mean age 22-23	21 texts	Short stories; mean 881 words, 2 columns.	16 misspellings and 14 grammar errors	3 min	S&P - Errors indicated verbally
Köpper (2016) E3	62-65	Most university students; mean age 24	14 texts	Short stories; mean 870 words, 2 columns.	16 misspellings and 14 grammar errors	3 min	S&P - Errors indicated verbally
Köpper (2016) E4	62-69	Most university students; mean age 24	20 texts	Short stories; mean 539 words.	12 misspellings and 8 grammar errors	2 min	S&P - Errors indicated verbally
Shibata (2015) E1	24	Mean age 31	2 texts	Newspaper articles (no current events, 'metaphorical or 'prosy' texts); Each 1 page, mean 660 chars.	Five sematic errors	4 min	S&P - Errors indicated verbally; annotations allowed
Wharton-Michael (2008)	42	University students, most in journalism	2 texts[13]	One newspaper article on smoking and one on music industry; each approx. 550 words.	17 proofing errors (grammar, spelling, punctuation, formatting, repeated words)	8 min	P – Errors corrected on paper S – Errors corrected on screen (track changes)
Gujar (1998) E1	10	-	1 text	Articles on animals from encyclopaedia. Each 1-page long, approx. 850 words.	20 word substitution errors (rhyming subs such as cake → fake)	-	S&P - Errors indicated verbally

11 N refers to the number per display condition (i.e., screen or paper).

12 Graduate Record Examination Comprehension section practice website.

13 'N' combines different non-display conditions (e.g., different fonts), which were included to test for interactions of other variables with display mode.

Experiment (E)	Participants		Materials proofread			Proofreading task	
	N[11]	Characteristics	N[11]	Characteristics	Errors per text	Time per text	Task
Ziefle (1998) E1	20	University students	3 texts[14]	Popular science texts from art, psychology and travel magazines; 7 pages long, approx. 260 words each page.	1 spelling error per around 120 words.	-	S&P - Errors indicated verbally
Gould (1987b) E1	18	From Yorktown Research Lab; age 20-60	3 texts[15]	Newspaper and magazine articles; each 5 pages long, approx. 1100 words.	Around 6-9 spelling errors.	-	S&P - Errors indicated verbally
Gould (1987b) E2	16	From Yorktown Research Lab; age 20-60	2 texts[15]	College texts on government and literature; 1000-word articles across multiple pages.	Around 6-9 spelling errors.	-	S&P - Errors indicated verbally
Gould (1987b) E3	12	From Yorktown Research Lab; age 20-50, mean 33	2 texts[15]	College texts on government and literature; 1000-word articles across multiple pages.	Around 6-9 spelling errors.	-	S&P - Errors indicated verbally
Gould (1987b) E4	15	From Yorktown Research Lab; age 18-48, mean 26	1 text[15]	College texts on government and literature; 1000-word articles across multiple pages.	Around 6-9 spelling errors.	-	S&P - Errors indicated verbally
Gould (1987b) E5	15	From Yorktown Research Lab; age 23-60, mean 36	1 text[14]	Newspaper and magazine articles; 1000 words each.	Around 6-9 spelling errors.	-	S&P - Errors indicated verbally
Gould (1987b) E6	18	From Yorktown Research Lab; age 18-48, mean 36	1 text14[15]	Newspaper and magazine articles, approx. 1100 words, multiple pages.	Around 6-9 spelling errors.	-	S&P - Errors indicated verbally
Gould (1987a) E2	12	Half with no CRT experience; half were members of Yorktown lab (daily CRT users)	1 text[15]	Newspaper and magazine articles, multiple pages.	1 spelling error per around 150 words.	10 min	S&P - Errors indicated verbally
Gould (1987a) E3	9	Research professionals and clerical people from the Yorktown lab	1 text[15]	Newspaper and magazine articles, multiple pages; each page about 235 words.	1 spelling error per around 150 words.	10 min	S&P - Errors indicated verbally

14 More than one screen type was compared in this study (e.g., screens with different resolutions); this number refers to the texts proofread for each screen type.
15 More than one paper type was compared in this study (e.g., good and poor quality paper); this number refers to the texts proofread for each paper type.

Experiment (E)	Participants		Materials proofread			Proofreading task	
	N[11]	Characteristics	N[11]	Characteristics	Errors per text	Time per text	Task
Gould (1987a) E6	6	Staff at IBM Human Factors Centre	1 text	Newspaper and magazine articles; 10-pages each.	1 spelling error per around 150 words.	10 min	S - P -
Gould (1987a) E7	9	Professionals and clerical people from Yorktown lab	1 text[15]	Newspaper and magazine articles; 10-pages each, about 230 words per page.	1 spelling error per around 150 words.	10 min	S&P - Errors indicated verbally
Gould (1987a) E8	10	Research and administrative staff from the IBM Yorktown lab	1 text[14]	Newspaper and magazine articles, 10-pages long	1 spelling error per around 150 words.	10 min	S&P - Errors indicated verbally
Gould (1987a) E9	10	Volunteers from IBM Yorktown lab	1 text	Newspaper and magazine articles, 5-pages long, each page approx. 235 words	1 spelling error per around 150 words.	10 min	S&P - Errors indicated verbally
Wilkinson (1987)	24	22-65 years, mean age 46. Five had some proofreading experience.	2 texts	Non-fiction text about the countryside; multiple pages.	2.13 errors per block (16 lines of 75 char.) of text: missing / additional spaces, misspellings, capitalisations.	50 min	S&P - Errors indicated verbally
Creed (1987) E1	30	Undergraduate students	1 text	Texts of clear prose and common vocabulary. Each text long enough to fill three VDU screens (3 x 25 lines of 80 chars).	18 errors: visually similar and dissimilar, letter substitutions, and syntactic errors.	Read once	P – Errors marked on paper S – Error marked on screen via joystick
Creed (1987) E2	24	Undergraduate students	2 texts	Texts of clear prose and common vocabulary. Each text long enough to fill 3 VDU screens (3 x 25 lines).	18 errors: visually similar and dissimilar, letter substitutions, and syntactic errors.	Read once	P – Errors marked on paper S – Error marked on screen via joystick
Gould (1984)	24	Clerk-typists; two age groups, mean ages 23 and 48	6 texts[15]	Articles that had more pages than could be read in the time period.	1 spelling error per around 150 words.	45 min	P – Errors circled S – Errors pointed to
Wright (1983)	16	Psychology research participant panel; mean age 45	2 texts	Articles from published sources, the "Observer" and "Punch". Text around 1500 words, 6 pages.	25 simple spelling and typing mistakes and 14 text manipulation errors.	Self-paced ~15 min	P – Type of error noted on paper S – Half the group noted the error type on screen and half on a separate paper

Table A - 2. Features of the screens and paper used in the empirical studies: screen type, and display size, resolution and positioning.

	Type	Size		Resolution		Position/movability	
Study	**Screen**	**Screen**	**Paper**	**Screen**	**Paper**	**Screen**	**Paper**
Hargis (2017)	-	-	-	-	-	-	-
Köpper (2016) E1	TFT-LCD Apple MacBook Pro	15.4”	21×30cm	128ppi (1680×1050)	600 dpi	Fixed - 75° backwards	Fixed - 15° (desk stand)
Köpper (2016) E2	TFT-LCD Apple MacBook Pro	15.4”	21×30cm	128ppi (1680×1050)	600 dpi	Fixed - 75° backwards	Fixed - 15° (desk stand)
Köpper (2016) E3	TFT-LCD Apple MacBook Pro	15.4”	21×30cm	128ppi (1680×1050)	600 dpi	Fixed - 75° backwards	Fixed - 15° (desk stand)
Köpper (2016) E4	Apple iPad 2 tablet	9.7”	15×21cm	132ppi (1024×768)	600 dpi	Fixed - 15°	Fixed - 15° (desk stand)
Shibata (2015) E1	Apple iPad tablet	-	B5	-	-	Moveable	Moveable
Wharton-Michael (2008)	-	-	-	-	-	-	-
Gujar (1998) E1	(1) dpiX monochrome (2) CRT NEC colour multi-synch monitor, 60hz	(1) - (2) 21”	8.5”x11”	(1) 282dpi (2) 85dpi	300dpi	Fixed - 15° from vertical	(1) Fixed - dpiX casing (2) Fixed - under glass in dpiX casing (3) Fixed - overhead transparency in dpiX casing (4) Moveable - typical paper
Ziefle (1998) E1	CRT monochrome antireflection (Sigma L-View)	19”	-	- 60 dpi (832×600) -120 dpi (1664×1200)	255dpi	-	Fixed (participants asked not to move the pages)
Gould (1987b) E1	CRT Mitsubishi colour C-3419C	32.5cm	-	480x640	240dpi	-	Moveable
Gould (1987b) E2	CRT Monochrome IBM 5080, 50Hz	12x12"	-	1024x1024	240dpi	-	Moveable
Gould (1987b) E3	CRT Monochrome IBM 5080, 60Hz	12x12"	-	1024x1024	240dpi	-	Moveable
Gould (1987b) E4	CRT Monochrome IBM 5080, 60Hz	12x12"	-	1024x1024	240dpi	-	Moveable
Gould (1987b) E5	CRT Moniterm Corp.	-	-	1024x768	240dpi	-	Moveable

Study	Type	Size		Resolution		Position/movability	
	Screen	**Screen**	**Paper**	**Screen**	**Paper**	**Screen**	**Paper**
Gould (1987b) E6	(1) CRT Moniterm Corp. (2) IBM Monochrome PC 5151 (3) IBM 3278	(1)12x9" (2) - (3) -	-	(1)1024x800 (2) - (3) -	Good quality: 240dpi Poor quality: -	-	Moveable
Gould (1987a) E2	CRT IBM 3277	-	-	-	-	-	Fixed - horizontal on table or vertical in copy holder
Gould (1987a) E3	CRT IBM 3277	-	-	-	-	-	Moveable
Gould (1987a) E6	CRT IBM 3277	-		-	-	-	Moveable
Gould (1987a) E7	CRT IBM 3277	-	-	-	-	-	Moveable
Gould (1987a) E8	(1) CRT IBM 3277 (2) CRT IBM 3278 (3) IBM 3290 gas-panel	-	-	-	-	-	Moveable
Gould (1987a) E9	CRT IBM 3277	-	(1)160 x 128 mm (2) 225 x180mm	-	-	-	-
Wilkinson (1987)	Teco Model TM1265	12”	A4	Centre res. of 800 lines at 20fL	-	-	-
Creed (1987) E1	Zenith Data Systems monochrome, green phosphor	12x10”	-	-	-	Fixed	Fixed - in wooden frame
Creed (1987) E2	Zenith Data Systems monochrome, green phosphor	12x10”	-	-	-	Fixed	Fixed - in wooden frame
Gould (1984)	CRT IBM 3277	-	-	-	-	-	Moveable
Wright (1983)	Apple II microcomputer black and white monitor	12”	-	-	-	-	-

Table A - 3. Comparison of polarity and text font types used in paper and screen conditions in empirical studies reviewed

Study	Screen type (provided for reference)	Polarity		Font	
		Screen	Paper	Screen	Paper
Hargis (2017)	-	-	-	-	-
Köpper (2016) E1	Apple MacBook Pro	Black on white	Black on white	12pt Helvetica	12pt Helvetica
Köpper (2016) E2	Apple MacBook Pro	Black on white	Black on white	12pt Helvetica	12pt Helvetica
Köpper (2016) E3	Apple MacBook Pro	Black on white	Black on white	12pt Helvetica	12pt Helvetica
Köpper (2016) E4	Apple iPad 2 tablet	Black on white	Black on white	12pt Helvetica	12pt Helvetica
Shibata (2015) E1	Apple iPad tablet	Black on white	Black on white	14pt	14pt
Wharton-Michael (2008)	-	-	-	-	-
Gujar (1998) E1	(1) dpiX monochrome; (2) CRT NEC colour	-	-	10pt Times Roman	10pt Times Roman
Ziefle (1998) E1	CRT monochrome	Dark on light-grey	Dark on light-grey	60dpi: Helvetica 12; 120dpi: Helvetica 24	Helvetica 12
Gould (1987b) E1	CRT Mitsubishi colour	Dark on white(ish)	Dark on white(ish)	12pt Letter Gothic; 12pt Press; 12pt Univers	12pt Letter Gothic; 12pt Press; 12pt Univers
Gould (1987b) E2	CRT IBM 5080	Dark on white(ish)	Dark on white(ish)	12pt Univers-65	12pt Univers-65
Gould (1987b) E3	CRT IBM 5080	Dark on white(ish)	Dark on white(ish)	12pt Univers-65	12pt Univers-65
Gould (1987b) E4	CRT IBM 5080	- Dark on light - Light on dark	Dark on white	12pt Univers-65	12pt Univers-65
Gould (1987b) E5	CRT Moniterm Corp.	Dark on white(ish)	Dark on white(ish)	Press-14	Press-14
Gould (1987b) E6	(1) CRT Moniterm; (2) IBM PC 5151; (3) IBM 3278	(1) Dark on whitish; (2)(3) Green on dark	Dark on white	(1) Press-14; (2) PC char. font; (3) 3278 char. font	Press-14
Gould (1987a) E2	CRT IBM 3277	Green on dark	-	7 x 9 matrix	10pt Letter Gothic
Gould (1987a) E3	CRT IBM 3277	Green on dark	-	7 x 9 matrix	10pt Letter Gothic
Gould (1987a) E6	CRT IBM 3277	Green on dark	-	7 x 9 matrix	10pt Letter Gothic
Gould (1987a) E7	CRT IBM 3277	Green on dark	i. Printed - ii. Colour photo of CRT green on dark	7 x 9 matrix	10pt Letter Gothic
Gould (1987a) E8	(1) CRT IBM 3277; (2) CRT IBM 3278; (3) IBM 3290 gas-panel	(1)(2) Green on dark; (3) Red on dark	Black on white	(1) 7x9 matrix; (2) 7x10 matrix; (3i) 5x10 matrix; (3ii) 7x12 matrix	10pt Letter Gothic
Gould (1987a) E9	CRT IBM 3277;	Green on dark	(1) dark on light (2) light on dark	7x9 matrix;	10pt Letter Gothic

Study	**Screen type** (provided for reference)	Polarity		Font	
		Screen	**Paper**	**Screen**	**Paper**
Wilkinson (1987)	Teco Model TM1265	Light green on darker green	White paper	-	Diablo Courier 10
Creed (1987) E1	Zenith monochrome, green phosphor	Light on dark	i. Printed - light on dark ii. Colour photo of VDU	-	-
Creed (1987) E2	Zenith monochrome, green phosphor	Light on dark	Light on dark	-	-
Gould (1984)	CRT IBM 3277	Green on dark	Dark on light	3277 char. font	10pt Letter Gothic
Wright (1983)	Apple II black and white	White on black	-	80 character board	-

Appendix B – Paper vs screen effects on accuracy reported by the studies reviewed

Notes about the content of the tables

- To make it easier to find information, we have divided the results into three tables based on the publication date.
- “Higher mean” is based on the paper and screen means; if means were not reported, it was based on the authors’ commentary about their results.
- "Significant difference” refers to the statistical significance of the comparison being tested, based on the authors’ criteria for significance.
- ? indicates that the information was not reported in the published article.
- ~ indicates that the mean is an approximation, which, in all cases, is because it was estimated from a graph of the data.
- Dotted lines in the tables group together results that are connected with each other.

Table B - 1 Details of the accuracy findings reported by the studies published 2017-1998.

Study	Comparison	Accuracy measure	Paper mean	Screen mean	Better mean	Significant
Hargis (2017)	Paper vs screen	Errors detected	Text 1: 8.74 (sd 3.17) Text 2: 8.97 (sd 2.66)	?	Paper	Yes
Köpper (2016) E1	Paper vs screen	Errors detected - false alarms	~12	~12	Screen	No
Köpper (2016) E2	Paper vs screen	Errors detected - false alarms	~12	~12	Screen	No
Köpper (2016) E3	1. Paper vs screen with normal luminance 2. Paper vs screen with reduced luminance	Errors detected - false alarms	~11 ~11	~11 ~12	Screen Screen	No No
Köpper (2016) E4	Paper vs screen	Errors detected - false alarms	~8	~ 8	Screen	No
Shibata (2015) E1	Paper vs screen	Errors detected	~60%	~50%	Paper	Yes
Wharton-Michael (2008)	1. Main effect of display mode: paper vs screen	Errors detected	9.77 (se 0.38)	7.78 (se 0.38)	Paper	Yes
	2. Main effect of topic of text: Napster vs smoking					?
	3. Interaction of display mode x topic of text 3a. Paper vs screen difference on Napster text 3b. Paper vs screen difference on Smoking text		 9.01 (se 0.57) 10.54 (se 0.51)	 8.25 (se 0.51) 7.33 (se 0.55)	 Paper Paper	Yes No Yes
Gujar (1998) E1	Overall comparison of six displays: 4 paper and 2 screen	Error rate	?	?	?	No
Ziefle (1998) E1	1. Paper vs screen with higher resolution 2. Paper vs screen with lower resolution	Errors detected	78.1% 78.1%	73.2% 69.9%	Paper Paper	Yes Yes

Table B - 2 Details of the accuracy findings reported by the studies published in 1987 by Gould and colleagues.

Study	Comparison	Measure	Paper mean	Screen mean	Better mean	Significant
Gould (1987b) E1	1. Main effect of display mode: paper vs screen	Errors detected	69%	70%	Screen	No
	2. Main effect of font type: Press vs Univers vs Letter Gothic					No
	3. Interaction of display mode x font type					No
	3a. Paper vs screen difference with Letter Gothic		74%	75%	Screen	?
	3b. Paper vs screen difference with Press font		68%	67%	Paper	?
	3c. Paper vs screen difference with Univers font		64%	67%	Paper	?
Gould (1987b) E2	1. Main effect of display mode: paper vs screen	Errors detected	78%	73%	Paper	No
	2. Main effect of page length: 22-line vs 28-lines					?
	3. Interaction of display mode x page length					No
	3a. Paper vs screen difference on small page		77%	73%	Paper	?
	3b. Paper vs screen difference on large page		79%	73%	Paper	?
Gould (1987b) E3	1. Main effect of display mode: paper vs screen	Errors detected	81%	79%	Paper	No
	2. Main effect of page length: 22-line vs 28-lines					No
	3. Interaction of display mode x page length					No
	3a. Paper vs screen difference on small page		83%	81%	Paper	?
	3b. Paper vs screen difference on large page		79%	77%	Paper	?
Gould (1987b) E4	1. Overall comparison of five displays: 1 paper and 4 screen	Errors detected	75%	74% (overall)	Paper	No
	1a. Paper vs screen with fixed contrast and positive polarity		75%	74%	Paper	?
	1b. Paper vs screen with fixed contrast and negative polarity		75%	75%	Same	?
	1c. Paper vs screen with adjustable contrast and positive polarity		75%	72%	Paper	?
	1d. Paper vs screen with adjustable contrast and negative polarity		75%	75%	Same	?
Gould (1987b) E5	1. Overall comparison of 3 displays: 1 paper and 2 screen	Errors detected				No
	1a. Paper vs aliased screen		77%	65%	Paper	?
	1b. Paper vs anti-aliased screen		77%	73%	Paper	?
Gould (1987b) E6	1. Overall comparison of 5 displays: 2 paper and 3 screen	Errors detected				Yes
	1a. poor quality paper vs 3278 screen		42%	64%	Screen	Yes
	1b: poor quality paper vs PC screen		42%	63%	Screen	Yes
	1c. poor quality paper vs anti-aliased screen		42%	72%	Screen	Yes
	1d. good quality paper vs 3278 screen		69%	64%	Paper	No
	1e. good quality paper vs PC screen		69%	63%	Paper	No
	1f. good quality paper vs anti-aliased screen		69%	72%	Screen	No

Study	Comparison	Measure	Paper mean	Screen mean	Better mean	Significant
Gould (1987a) E2	1. Overall comparison of 3 displays: 2 paper and 1 screen 1a. Paper-horizontal vs screen 1b. Paper-vertical vs screen	Errors detected	 61% 67%	 65% 65%	 Screen Paper	No ? ?
Gould (1987a) E3	1. Overall comparison of 3 displays: 2 paper and 1 screen 1a. Paper-normal vs screen 1b. Paper-rotated vs screen	Errors detected	 76% 76%	 75% 75%	 Paper Paper	No ? ?
Gould (1987a) E7	1. Overall comparison of 3 displays: 2 paper and 1 screen 1a. printed paper vs screen 1b. photo of screen vs screen	Errors detected	 85% 82%	 69% 69%	 Paper Paper	Yes Yes Yes
Gould (1987a) E8	1. Overall comparison of 5 displays: 1 paper and 4 screen 1a. Paper vs 3277 screen 1b. Paper vs 3278 screen 1c. Paper vs 3290 small-page size screen 1d. Paper vs 3290 large-page size screen	Errors detected	 84% 84% 84% 84%	 85% 88% 89% 85%	 Screen Screen Screen Screen	No ? ? ? ?
Gould (1987a) E9	1. Overall comparison of 5 displays: 4 paper and 1 screen 1a. Paper with positive polarity and paper aspect ratio vs screen 1b. Paper with positive polarity and screen aspect ratio vs screen 1c. Paper with negative polarity and paper aspect ratio vs screen 1d. Paper with negative polarity and screen aspect ratio vs screen	Errors detected	 77% 76% 79% 78%	 90% 90% 90% 90%	 Screen Screen Screen Screen	No ? ? ? ?

Table B - 3 Details of the accuracy findings reported by studies published in 1987-1983 (excluding Gould's studies in Table B - 2).

Study	Comparison	Measure	Paper mean	Screen mean	Better mean	Significant
Wilkinson (1987)	1. Main effect of display mode: paper vs screen	Errors missed	?	?	?	Yes
	2. Main effect of 10-min period of test					Yes
	3. Interaction between display effect and time period of test					Yes
	3a. Paper vs screen in 1st 10min-period		~22%	~25%	Paper	?
	3b. Paper vs screen in 2nd 10min-period		~25%	~33%	Paper	?
	3c. Paper vs screen in 3rd 10min-period		~25%	~30%	Paper	?
	3d. Paper vs screen in 4th 10min-period		~22%	~30%	Paper	?
	3e. Paper vs screen in 5th 10min-period		~25%	~38%	Paper	?
Creed (1987) E1	1. Main effect of display: 2 paper and 1 screen condition	Error detection				Yes
	1a. Printed paper vs screen		4.8	4.3	Paper	Yes
	1b. Photo of screen (paper condition) vs screen		4.5	4.3	Paper	No
	2. Main effect of error type					Yes
	3. Interaction of display mode x error type					No
	3a. Printed paper vs screen on visually similar errors		4.8	3.8	Paper	?
	3b. Printed paper vs screen on syntactic errors		4.4	4.2	Paper	?
	3c. Printed paper vs screen on visually dissimilar errors		5.1	4.9	Paper	?
	3d. Photo vs screen on visually similar errors		4.0	3.8	Paper	?
	3e. Photo vs screen on syntactic errors		4.4	4.2	Paper	?
	3f. Photo vs screen on visually dissimilar errors		5.0	4.9	Paper	?
Creed (1987) E2	1. Main effect of display mode: paper vs screen	Error detection	?	?	?	Yes
	2. Main effect of error type: visually similar, syntactic, visual dissimilar		?	?	?	Yes
	3. Main effect of column format: column vs single column		?	?	?	No
	4. Interaction between display mode and error type					Yes
	4a. Paper vs screen with visually similar errors		?	?	?	?
	4b. Paper vs screen with syntactic errors		?	?	?	?
	4c. Paper vs screen with visually dissimilar errors		?	?	?	?
	5. Interaction between display mode and column format					No
	5a. Paper vs screen with column format		4.7	4.4	Paper	Yes
	5b. Paper vs screen with single-column format		4.7	4.0	Paper	Yes
	6. Interaction between display mode x error type x column format					No
	6a. Paper vs screen with column format and visually similar errors		4.2	3.3	Paper	?
	6b. Paper vs screen with column format and syntactic errors		4.6	3.6	Paper	?
	6c. Paper vs screen with column format and visually dissimilar errors		5.3	5.0	Paper	?

Study	Comparison	Measure	Paper mean	Screen mean	Better mean	Significant
	6d. Paper vs screen with single column and visually similar errors		4.2	3.9	Paper	?
	6e. Paper vs screen with single column and syntactic errors		4.9	4.1	Paper	?
	6c. Paper vs screen with single column and visually dissimilar errors		5.0	5.1	Screen	?
Gould (1984)	1. Main effect of display mode: paper vs screen	Errors missed	30%	33%	Paper	No
	2. Main effect of work period: 6 x 1 hour work periods					?
	3. Interaction between display effect and work period					No
	3a. Paper vs screen in 1st hour		~30%	~30%	Paper	?
	3b. Paper vs screen in 2nd hour		~30%	~30%	Paper	?
	3c. Paper vs screen in 3rd hour		~30%	~35%	Paper	?
	3d. Paper vs screen in 4th hour		~30%	~35%	Paper	?
	3e. Paper vs screen in 5th hour		~30%	~30%	Paper	?
	3f. Paper vs screen in 6th hour		~30%	~35%	Paper	?
Wright (1983)	1a. Main effect of display mode: paper vs screen (screen group annotating on screen)	Errors missed	6.75	7.54	Paper	No
	1b. Main effect of text: 1st vs 2nd text (screen group annotating on screen)					No
	1c. Interaction between display mode and text (screen group annotating on screen)					?
	1ci. Paper vs screen for text 1 (screen group annotating on screen)		7.31	7.63	Paper	?
	1cii. Paper vs screen for text 2 (screen group annotating on screen)		6.19	7.44	Paper	?
	2a. Main effect of display mode: paper vs screen (for screen group annotating on paper)		7.19	8.13	Paper	Yes
	2b. Main effect of text: 1st vs 2nd text (screen group annotating on paper)					No
	2c. Interaction between display mode and text (screen group annotating on paper)					?
	2ci. Paper vs screen for text 1 (screen group annotating on paper)		7.55	8.63	Paper	?
	2cii. Paper vs screen for text 2 (screen group annotating on paper)		6.63	7.63	Paper	?

Appendix C – Paper vs screen effects on speed reported by the studies reviewed

Notes about the content of the tables

- To make it easier to find information, we have divided the results into three tables based on the publication date.
- "Faster mean" is based on the paper and screen means; if means were not reported, it was based on the authors' commentary about their results.
- "Significant difference" refers to the statistical significance of the comparison being tested, based on the authors' criteria for significance.
- ? indicates that the information was not reported in the published article.
- ~ indicates that the mean is an approximation, which, in all cases, is because it was estimated from a graph of the data.
- Dotted lines in the tables group together results that are connected with each other.

Table C - 1 Details of the speed findings reported by studies published 2017-1998.

Study	Comparison	Measure	Paper mean	Screen mean	Faster mean	Significant
Köpper (2016) E1	Paper vs screen	Number of words read	~500	~500	Screen	No
Köpper (2016) E2	Paper vs screen	Number of words read	~ 525	~ 525	Paper	No
Köpper (2016) E3	1. Paper vs screen with normal luminance 2. Paper vs screen with reduced luminance	Number of words read	~ 525 ~ 525	~ 525 ~ 500	Screen Paper	No No
Köpper (2016) E4	Paper vs screen	Number of words read	~ 300	~ 325	Screen	Yes
Shibata (2015) E1	Paper vs screen	Time to detect all five errors (seconds)	218.5	225.4	Paper	No
Gujar (1998) E1	Overall comparison of six displays: 4 paper and 2 screen	Response time	?	?	?	No
Ziefle (1998) E1	1. Paper vs screen with higher resolution 2. Paper vs screen with lower resolution	Reading rate (words/min)	201 201	182 179	Paper Paper	Yes Yes

Table C - 2 Details of the speed findings reported by the studies published in 1987 by Gould and colleagues.

Study	Comparison	Measure	Paper mean	Screen mean	Faster mean	Significant
Gould (1987b) E1	1. Main effect of display mode: paper vs screen	Reading rate (words/min)	220	209	Paper	Yes
	2. Main effect of font type: Press vs Univers vs Letter Gothic					No
	3. Interaction of display mode x font type					No
	3a. Paper vs screen difference with Letter Gothic		217	202	Paper	No
	3b. Paper vs screen difference with Press font		226	212	Paper	No
	3c. Paper vs screen difference with Univers font		219	213	Paper	No
Gould (1987b) E2	1. Main effect of display mode: paper vs screen	Reading rate (words/min)	201	196	Paper	No
	2. Main effect of page length: 22-line vs 28-lines					No
	3. Interaction of display mode x page length					No
	3a. Paper vs screen difference on small page		198	190	Paper	?
	3b. Paper vs screen difference on large page		205	203	Paper	?
Gould (1987b) E3	1. Main effect of display mode: paper vs screen	Reading rate (words/min)	206	204	Paper	No
	2. Main effect of page length: 22-line vs 28-lines					No
	3. Interaction of display mode x page length					No
	3a. Paper vs screen difference on small page		200	204	Screen	?
	3b. Paper vs screen difference on large page		212	204	Paper	?
Gould (1987b) E4	1. Overall comparison of five displays: 1 paper and 4 screen	Reading rate (words/min)	252	241 overall	Paper	No
	1a. Paper vs screen with fixed contrast and positive polarity		252	254	Screen	?
	1b. Paper vs screen with fixed contrast and negative polarity		252	236	Paper	?
	1c. Paper vs screen with adjustable contrast and positive polarity		252	240	Paper	?
	1d. Paper vs screen with adjustable contrast and negative polarity		252	233	Paper	?
Gould (1987b) E5	1. Overall comparison of 3 displays: 1 paper and 2 screen	Reading rate (words/min)				Yes
	1a. Paper vs aliased screen		262	240	Paper	Yes
	1b. Paper vs anti-aliased screen		262	252	Paper	No
Gould (1987b) E6	1. Overall comparison of 5 displays: 2 paper and 3 screen	Reading rate (words/min)				Yes
	1a. poor paper vs 3278 screen		175	189	Screen	Yes
	1b: poor paper vs PC screen		175	207	Screen	Yes
	1c. poor paper vs anti-aliased screen		175	218	Screen	Yes
	1d. good paper vs 3278 screen		236	189	Paper	Yes
	1e. good paper vs PC screen		236	207	Paper	Yes
	1f. good paper vs anti-aliased screen		236	218	Paper	No
Gould (1987a) E2	1. Overall comparison of 3 displays: 2 paper and 1 screen	Reading rate (words/min)				Yes
	1a. Paper-horizontal vs screen		216	185	Paper	Yes

Study	Comparison	Measure	Paper mean	Screen mean	Faster mean	Significant
	1b. Paper-vertical vs screen		216	185	Paper	Yes
Gould (1987a) E3	1. Overall comparison of 3 displays: 2 paper and 1 screen	Reading rate (words/min)				Yes
	1a. Paper-normal vs screen		206	184	Paper	Yes
	1b. Paper-rotated vs screen		211	184	Paper	Yes
Gould (1987a) E6	Paper vs screen	Reading rate (words/min)	200	149	Paper	Yes
Gould (1987a) E7	1. Overall comparison of 3 display: 2 paper and 1 screen	Reading rate (words/min)				Yes
	1a. printed paper vs screen		222	165	Paper	Yes
	1b. photo of screen vs screen		176	165	Paper	Yes
Gould (1987a) E8	1. Overall comparison of 5 displays: 1 paper and 4 screen	Reading rate (words/min)				Yes
	1a. Paper vs 3277 screen		200	167 words	Paper	Yes
	1b. Paper vs 3278 screen		200	172 words	Paper	Yes
	1c. Paper vs 3290 small-page size screen		200	170 words	Paper	Yes
	1d. Paper vs 3290 large-page size screen		200	172 words	Paper	Yes
Gould (1987a) E9	1. Overall comparison of 5 display: 4 paper and 1 screen	Reading rate (words/min)				No
	1a. Paper with positive polarity and paper aspect ratio vs screen		205	198 words	Paper	?
	1b. Paper with positive polarity and screen aspect ratio vs screen		208	198 words	Paper	?
	1c. Paper with negative polarity and paper aspect ratio vs screen		206	198 words	Paper	?
	1d. Paper with negative polarity and screen aspect ratio vs screen		215	198 words	Paper	?

Table C - 3 Details of the speed findings reported by studies published in 1987-1983 (excluding Gould’s studies in Table C - 2).

Study	Comparison	Measure	Paper mean	Screen mean	Faster mean	Significant
Wilkinson (1987)	1. Main effect of display mode	Lines read in 50 minute task	?	?	?	Yes
	2. Main effect of 10-min period of test		?	?	?	Yes
	3. Interaction between display effect and time period of test					No
	3a. Paper vs screen in 1st 10min-period		~145	~110	Paper	?
	3b. Paper vs screen in 2nd 10min-period		~150	~115	Paper	?
	3c. Paper vs screen in 3rd 10min-period		~150	~115	Paper	?
	3d. Paper vs screen in 4th 10min-period		~150	~115	Paper	?
	3e. Paper vs screen in 5th 10min-period		~150	~125	Paper	?
Creed (1987) E1	1. Main effect of display: 2 paper and 1 screen condition	Performance time (seconds)				No
	1a. Printed paper vs screen		356	366	Paper	?
	1b. Photo of screen (paper condition) vs screen		368	366	Screen	?
	2. Main effect of error type					No
	3. Interaction of display mode x error type					?
	3a. Printed paper vs screen on visually similar errors		?	?	?	?

Study	Comparison	Measure	Paper mean	Screen mean	Faster mean	Significant
	3b. Printed paper vs screen on syntactic errors		?	?	?	?
	3c. Printed paper vs screen on visually dissimilar errors		?	?	?	?
	3d. Photo vs screen on visually similar errors		?	?	?	?
	3e. Photo vs screen on syntactic errors		?	?	?	?
	3f. Photo vs screen on visually dissimilar errors		?	?	?	?
Creed (1987) E2	1. Main effect of display mode: paper vs screen	Performance time (seconds)	?	?	?	Yes
	2. Main effect of error type		?	?	?	No
	3. Main effect of column format		?	?	?	No
	4. Interaction between display mode and error type		?	?	?	No
	4a. Paper vs screen with visually similar errors		?	?	?	?
	4b. Paper vs screen with syntactic errors		?	?	?	?
	4c. Paper vs screen with visually dissimilar errors		?	?	?	?
	5. Interaction between display mode and column format					No
	5a. Paper vs screen with column format		347	379	Paper	Yes
	5b. Paper vs screen with single-column format		369	400	Paper	Yes
	6. Interaction between display mode x error type x column format					No
	6a. Paper vs screen with column format and visually similar errors		?	?	?	?
	6b. Paper vs screen with column format and syntactic errors		?	?	?	?
	6c. Paper vs screen with column format and visually dissimilar errors		?	?	?	?
	6d. Paper vs screen with single column and visually similar errors		?	?	?	?
	6e. Paper vs screen with single column and syntactic errors		?	?	?	?
	6c. Paper vs screen with single column and visually dissimilar errors		?	?	?	?
Gould (1984)	1. Main effect of display mode: paper vs screen	Speed (pages/min; ~words)	0.99; ~205	0.77; ~159	Paper	Yes
	2. Main effect of work period: 6 x 1 hour work periods		?	?	?	Yes
	3. Interaction between display effect and work period					No
	3a. Paper vs screen in 1st hour		~0.8	~0.7	Paper	?
	3b. Paper vs screen in 2nd hour		~0.8	~0.7	Paper	?
	3c. Paper vs screen in 3rd hour		~1	~0.8	Paper	?
	3d. Paper vs screen in 4th hour		~1	~0.7	Paper	?
	3e. Paper vs screen in 5th hour		~1	~0.8	Paper	?
	3f. Paper vs screen in 6th hour		~1	~0.8	Paper	?
Wright (1983)	1a. Main effect of display mode: paper vs screen (for screen group annotating on screen)	Mean time to proofread text (minutes)	?	?	Paper	Yes
	1b. Main effect of text: 1st vs 2nd (for screen group annotating on screen)					Yes
	1c. Interaction between display mode and text (for screen group annotating on screen)					Yes

Study	Comparison	Measure	Paper mean	Screen mean	Faster mean	Significant
	1ci. Paper vs screen for 1st text (for screen group annotating on screen)		11.0	16.0	Paper	?
	1cii. Paper vs screen for 2nd text (for screen group annotating on screen)		9.9	12.8	Paper	?
	2a. Main effect of display mode: paper vs screen (for screen group annotating on paper)		?	?	Paper	Yes
	2b. Main effect of text: 1st vs 2nd (for screen group annotating on paper)					Yes
	2c. Interaction between display mode and text (for screen group annotating on paper)					No
	2ci. Paper vs screen for 1st text (for screen group annotating on paper)		10.2	13.4	Paper	?
	2cii. Paper vs screen for 2nd text (for screen group annotating on paper)		9.1	12.0	Paper	?