Lesen am Bildschirm

Ein systematischer Vergleich von TFT-LCDs mit Papier und der Einfluss hoher Bildschirmauflösungen auf objektive Leistungsmaße und subjektives Wohlbefinden

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Maja Köpper

aus Düsseldorf

Düsseldorf, August 2015

Aus dem Institut für Experimentelle Psychologie

der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit Genehmigung

der Mathematisch-Naturwissenschaftlichen Fakultät der

Heinrich-Heine-Universität Düsseldorf

Referent: Prof. Dr. Susanne Mayr

Koreferent: Prof. Dr. Axel Buchner

Tag der mündlichen Prüfung:

Inhaltsverzeichnis

Zusammenfassung

Frühe Studien aus den 80er und 90er Jahren des letzten Jahrhunderts konnten einen deutlichen Leistungsnachteil sowie Einschränkungen des physischen Wohlbefindens beim Lesen am Bildschirm im Vergleich zum Lesen auf Papier nachweisen. Heutige Bildschirme zeichnen sich durch eine deutlich gesteigerte Bildschirmqualität aus, so dass sich die Frage stellt, ob die Nachteile des Lesens am Bildschirm auch heute noch existieren und welche Leistungsverbesserungen durch die neuesten Weiterentwicklungen der Bildschirmqualität zu erwarten sind. Im ersten Teil der Arbeit wurde in insgesamt vier Experimenten das Lesen an einem modernen Bildschirm mit dem Lesen auf Papier verglichen. Bei einer Lesedauer von bis zu einer Stunde zeigten sich keine Unterschiede in der Lesegeschwindigkeit und der Korrekturleseleistung, die Teilnehmer berichteten jedoch von stärkeren Augenbeschwerden beim Lesen am Bildschirm und äußerten eine deutliche Präferenz für das Lesen auf Papier. Eine zu starke Hintergrundhelligkeit des Bildschirms konnte als Ursache für die unterschiedlich stark ausgeprägten Augenbeschwerden ausgeschlossen werden. Durch die Positionierung beider Medien im gleichen Neigungswinkel von 15° konnte dieser Unterschied in den Augenbeschwerden jedoch eliminiert werden. Eine papierähnliche Bildschirmpositionierung scheint folglich unbeeinträchtigtes Lesen am Bildschirm ohne negative physische Auswirkungen zu ermöglichen. Ausgehend von der Äquivalenz der Medien hinsichtlich Lesegeschwindigkeit und Korrekturleseleistung sollte im zweiten Teil der Arbeit mittels zweier Experimente überprüft werden, ob eine Verbesserung der Darstellungsqualität am Bildschirm durch die Verwendung einer sehr hohen Bildschirmauflösung zu einer Leistungssteigerung beim Lesen am Bildschirm führen kann. Es zeigte sich kein Einfluss der Bildschirmauflösung auf das Leseverständnis, die Lesezeit, die Korrekturlesegeschwindigkeit und die Korrekturleseleistung beim Vergleich eines 132 ppi Bildschirms mit einem 264 ppi Bildschirm. Die subjektiven Einschätzungen zum physischen Wohlbefinden ergaben jedoch signifikant stärker ausgeprägte Kopfschmerzsymptome und muskuläre Beschwerden in der 132 ppi Bedingung im Vergleich zur 264 ppi Bedingung. Um physische Beeinträchtigungen zu vermeiden, ist die Verwendung hochauflösender Bildschirme folglich zu empfehlen.

Abstract

Early studies have shown considerable disadvantages in performance and physical well-being when reading text from screen as compared to paper. Considering the advanced display quality of modern screens it is not clear if the reported disadvantages in reading on screen can still be found today or to what extent you can expect improvements in reading performance. In the first part of this work reading on a modern computer screen was compared with reading on paper in a series of four experiments. With a reading duration of up to an hour proofreading speed and performance were equal for a computer screen and a paper display, but participants reported stronger eyestrain symptoms in the screen condition and a strong preference for reading on paper. It could be excluded that the difference in eyestrain symptoms was caused by an excessively high luminance of the computer screen. Positioning both displays in equal inclination angles of 15° eliminated this difference. A paper-like positioning of a screen seems to enable unimpaired reading without evidence of physical strain. Given the equivalence of the reading devices with regard to reading speed and proofreading performance, two additional experiments were conducted to test whether the improvement of display quality by means of high display resolution has beneficial effects on visual performance. There was no effect of display resolution on reading comprehension, reading time, proofreading speed and performance between a 132 ppi and a 264 ppi display. However, subjective ratings of physical discomfort revealed significantly more complaints about headache and musculoskeletal strain in the 132 ppi condition than in the 264 ppi condition. Thus, in order to avoid physical impairments displays with high pixel densities are recommended.

1. Lesen am Bildschirm im Vergleich zum Lesen auf Papier: Gibt es noch einen Unterschied?

Die Arbeit am Computer und die Nutzung verschiedener Bildschirmtechnologien im Alltag ist für die meisten Menschen in industrialisierten Gesellschaften allgegenwärtig. Bereits im Jahr 2012 beinhalteten 86 % aller Arbeitsplätze in Deutschland Tätigkeiten am Computer (Statistisches Bundesamt, 2012a) und 81 % aller Haushalte verfügten über Computer, Laptops, Notebooks, Netbooks oder Tablet Computer (Statistisches Bundesamt, 2012b). Die starke Verbreitung von Computerbildschirmen ist vor allem auf die vielen Vorteile dieser Technologie zurückzuführen. Dazu zählen zum Beispiel die schnellere Verfügbarkeit und Übermittlung von Informationen, aber auch die digitale Datenspeicherung sowie die damit verbundene Platzersparnis. Es ist anzunehmen, dass der Anteil elektronisch verfügbarer Informationen, die damit einhergehende Nutzung verschiedener Bildschirmtechnologien und die sukzessive Verdrängung von Informationen auf Papier in den nächsten Jahren weiter zunehmen wird. Dabei stellt sich zunächst die Frage, ob ein Computerbildschirm die gleiche Lesbarkeit aufweist wie bedrucktes Papier und ob eventuelle Unterschiede zwischen diesen Medien Einbußen in Leistungsmaßen wie Lesegeschwindigkeit und Lesegenauigkeit oder Beeinträchtigungen des subjektiven Wohlbefindens beim Lesen am Bildschirm nach sich ziehen.

Mit Einführung der ersten Computer im Arbeits- und Privatbereich in den 1980er Jahren wurden zahlreiche Studien durchgeführt, die genau dieser Frage nachgingen. Es zeigte sich, dass die Lesegeschwindigkeit am Bildschirm um bis zu 20 - 30 % langsamer war als die Lesegeschwindigkeit beim Lesen auf Papier (Creed, Dennis, & Newstead, 1987 [Exp. 2]; Gould, Alfaro, Barnes, et al., 1987; Gould & Grischkowsky, 1984; Heppner, Anderson, Farstrup, & Weiderman, 1985; Muter, Latremouille, Treurniet, & Beam, 1982; Wilkinson & Robinshaw, 1987; P. Wright & Lickorish, 1983). Darüber hinaus konnte in einigen Studien belegt werden, dass auch die Korrekturlesegenauigkeit durch die Arbeit am Bildschirm beeinträchtigt wird (Creed et al., 1987; Wilkinson & Robinshaw, 1987; P. Wright & Lickorish, 1983). Neben diesen Leistungseinbußen zeigten sich auch nachteilige Effekte des Lesens am Bildschirm auf das subjektive Wohlbefinden. Die Hauptsymptome von Computernutzern waren Augenbeschwerden, wie zum Beispiel trockene oder brennende Augen (Cole, Maddocks, & Sharpe, 1996; Knave, Wibom, Voss, Hedström, & Bergqvist, 1985; Läubli, Hünting, & Grandjean, 1981; Ong, Hoong, & Phoon, 1981; Rey & Meyer, 1980; Sheedy, 1992; Smith, Cohen, Stammerjohn, & Happ, 1981) sowie Muskelbeschwerden im Nacken- und Schulterbereich und im Rücken (Bergqvist, Wolgast, Nilsson, & Voss, 1995; Cole et al., 1996; Hünting, Läubli, & Grandjean, 1981; Ong et al., 1981; Sheedy, 1992; Smith et al., 1981; Starr, 1983). Zudem zeigte sich eine deutliche Nutzerpräferenz für das Lesen auf Papier (z.B. Gould, Alfaro, Finn, Haupt, & Minuto, 1987; Heppner et al., 1985; Oborne & Holton, 1988).

Rückschlüsse auf mögliche Unterschiede zwischen dem Lesen am Bildschirm und dem Lesen auf Papier in der heutigen Zeit sind auf Grundlage der genannten Untersuchungen jedoch nicht ohne weiteres möglich. Zum einen sind sowohl die berichteten Leistungseinbußen als auch die Beeinträchtigung des subjektiven Wohlbefindens beim Lesen am Bildschirm größtenteils auf die Eigenschaften der verwendeten Bildschirmtechnologien zurückzuführen. Bis um die Jahrtausendwende waren CRT-Bildschirme (cathode ray tube) weit verbreitet. Um ein sichtbares Bild zu erzeugen, wird bei einem CRT-Bildschirm ein gebündelter Elektronenstrahl auf die Innenseite der Bildschirmoberfläche gelenkt und trifft dort auf eine fluoreszierende Phosphorbeschichtung, die so für eine begrenzte Zeit zum Leuchten gebracht wird. Der Elektronenstrahl aktiviert kontinuierlich Bildpunkt für Bildpunkt und Zeile für Zeile der gesamten Bildschirmoberfläche. Die Zeit, die benötigt wird, um die gesamte Oberfläche zu erleuchten, bestimmt die Bildwiederholfrequenz, die je nach Bildschirmmodell zwischen 60 und 100 Hz liegen kann. Zu geringe Bildwiederholfrequenzen in Kombination mit der begrenzten Leuchtzeit der verwendeten Phosphorbeschichtung können beim Betrachter einem Flimmereindruck hervorrufen. Dieses Flimmern des Bildschirms kann sich nachteilig auf die Lesegeschwindigkeit (Montegut, Bridgeman, & Sykes, 1997) sowie das Wohlbefinden des Lesers (Harwood & Foley, 1987; Läubli et al., 1981) auswirken.

Die CRT-Technologie wurde jedoch seit Beginn des 21. Jahrhunderts weitestgehend durch Flüssigkristallbildschirme, sogenannte TFT-LCDs (thin-film-transistor liquid-crystal-displays) ersetzt. Bei TFT-LCDs besteht jeder Bildpunkt aus einer oder mehreren mit Flüssigkristallen gefüllten Zellen, die einzeln angesteuert werden können. Je nachdem, ob Spannung an einer dieser Zellen angelegt ist, ändert sich die Ausrichtung der Flüssigkristalle, so dass die Lichtdurchlässigkeit jeder einzelnen Zelle direkt gesteuert werden kann. Somit ist bei TFT-LCDs, im Gegensatz zu CRT-Bildschirmen, eine flimmerfreie Darstellung von Texten am Bildschirm möglich. Ein weiterer Nachteil älterer CRT-Bildschirme besteht in der vergleichsweise geringen Bildschirmauflösung (Anzahl der Pixel pro Bildschirmhorizontale und -vertikale) bzw. Punktdichte (Anzahl der Pixel pro Inch [ppi]), im Bereich zwischen 60 und 90 ppi (z.B. Alt & Noda, 1998; Gould, Alfaro, Finn, et al., 1987; Ziefle, 1998). Eine zu geringe Auflösung eines CRT-Bildschirms kann zu Leistungseinbußen führen wie zum Beispiel einer Verlangsamung der Lesegeschwindigkeit bei Korrekturleseaufgaben (z.B. Gould, Alfaro, Finn, et al., 1987) sowie bei Aufgaben zur visuellen Suche (z.B. Ziefle, 1998). Höhere Bildschirmauflösungen werden darüber hinaus gegenüber geringen bevorzugt (Ziefle, 1998). Gängige TFT-LCDs verfügen über deutlich höhere Bildschirmauflösungen ab 120 ppi (Alt & Noda, 1998). Zudem ermöglichen TFT-LCDs im Vergleich zu älteren CRT-Bildschirmen höhere Hintergrundhelligkeiten und, damit einhergehend, höhere maximale Helligkeitskontraste (Menozzi, Lang, Näpflin, Zeller, & Krueger, 2001).

Die genannten technischen Einschränkungen älterer CRT-Bildschirme wurden in Zusammenhang gebracht mit den beschriebenen Leistungseinbußen sowie der Verschlechterung des subjektiven Wohlbefindens beim Lesen am Bildschirm im Vergleich zur gewohnten Textdarstellung auf Papier (z.B. Gould, Alfaro, Finn, et al., 1987; Oborne & Holton, 1988). In direkten Vergleichen von TFT-LCDs und CRT-Bildschirmen zeigte sich zudem eine verbesserte visuelle Leistung an TFT-LCDs zum Beispiel in Geschwindigkeit, Fehlerrate sowie Augenbewegungsparametern bei der visuellen Suche (Menozzi et al., 2001; Menozzi, Näpflin, & Krueger, 1999; Näsänen, Karlsson, & Ojanpää, 2001; Ziefle, 2001) und bei der Buchstabenidentifikation (Shieh & Lin, 2000). Darüber hinaus wurden TFT-LCDs gegenüber CRT-Bildschirmen bevorzugt (Chen & Lin, 2004; Shieh & Lin, 2000; Ziefle, 2001). Diese Studien beinhalteten jedoch keine Bedingung, in der die Teilnehmer Text auf Papier lesen sollten. Es wird also deutlich, dass die TFT-LCD-Technologie den CRT-Bildschirmen überlegen ist; die Frage, ob es Unterschiede zwischen dem Lesen auf TFT-LCDs und dem Lesen auf Papier gibt, ist jedoch weiterhin ungeklärt.

Neben den Verbesserungen der technischen Bildschirmeigenschaften gab es in den vergangenen Jahren entscheidende Entwicklungen in der Textdarstellung am Bildschirm. Zum einen wurde Text vor allem bei älteren CRT-Bildschirmen vorwiegend negativ polar dargestellt (z.B. weiße Schrift auf schwarzen Hintergrund). Im Vergleich zu der auf Papier üblichen positiv polaren Darstellung (z.B. schwarze Schrift auf weißem Hintergrund) führt eine negative Textpolarität jedoch zu einer Verschlechterung der Lesbarkeit (z.B. Buchner & Baumgartner, 2007; Buchner, Mayr, & Brandt, 2009; Mayr & Buchner, 2010; Piepenbrock, Mayr, Mund, & Buchner, 2013). Dementsprechend wird heute die Verwendung positiver Textpolarität auch bei der Darstellung am Bildschirm empfohlen. Durch die Einschränkungen früherer Textverarbeitungsprogramme war darüber hinaus die Textdarstellung auf Bildschirm und Papier hinsichtlich verschiedener Layoutvariablen, wie zum Beispiel Schriftart, Zeilenabstand oder Zeilenlänge, in vielen frühen Studien nicht äquivalent (z.B. Gould & Grischkowsky, 1984; Muter et al., 1982). Heutzutage ermöglichen gängige Textverarbeitungsprogramme eine verbesserte Textdarstellung am Bildschirm, die der Darstellung auf Papier entspricht, und erhöhen somit die Vergleichbarkeit beider Medien. Moderne Betriebssysteme verfügen zudem über verschiedene Schriftenglättungsalgorithmen zur optischen Verfeinerung des Schriftbildes. Bei diesen Techniken werden die einzelnen Bildpunkte unterschiedlich eingefärbt, um so unerwünschte Nebeneffekte einer unzureichenden Bildschirmauflösung, wie zum Beispiel Treppeneffekte bei der Darstellung diagonaler Linien in Buchstaben, zu vermeiden. Die Verwendung von Schriftenglättung durch Graustufen führt im Vergleich zu herkömmlicher Textdarstellung zu einer signifikant gesteigerten Lesegeschwindigkeit, geringeren physischen Beschwerden (Sheedy & McCarthy, 1994) sowie höherer subjektiver Lesbarkeit und Präferenz (Gould, Alfaro, Finn, et al., 1987). Neben der Verwendung von Graustufen kann bei TFT-LCDs durch die gezielte Ansteuerung einzelner farbiger Subpixel, dem sogenannten Subpixel-Rendering, die horizontale Auflösung eines TFT-LCDs um das Dreifache erhöht werden (z.B. Slattery & Rayner, 2010). Dies führt zum Beispiel zu Verbesserungen der Lesegeschwindigkeit (Dillon, Kleinman, Choi, & Bias, 2006; Slattery & Rayner, 2010), der subjektiven Lesbarkeit (Gugerty, Tyrrell, Aten, & Edmonds, 2004) sowie zu höherer Präferenz gegenüber Textdarstellungen ohne Schriftenglättung (Dillon et al., 2006; Gugerty et al., 2004).

Ein weiterer Faktor ist schließlich auch die Vertrautheit mit den verschiedenen Medien. Die Teilnehmer der genannten Studien waren mit dem Lesen am Computerbildschirm im Vergleich zum Lesen auf Papier weniger vertraut, als dies heute der Fall ist. Diese Tatsache liefert möglicherweise eine zusätzliche Erklärung für die genannten Leistungsunterschiede beim Lesen auf Papier im Vergleich zum Lesen an CRT-Bildschirmen sowie für die starke Präferenz für das Lesen auf Papier in dieser Zeit.

Die genannten Faktoren verdeutlichen, dass sich die Ergebnisse der frühen Untersuchungen nicht zwangsläufig auf die heutige Zeit übertragen lassen. Aktuelle Untersuchungen, die moderne TFT-LCDs mit Papier vergleichen, sind jedoch selten und die Befunde in Bezug auf mögliche Unterschiede zwischen den beiden Medien uneinheitlich. Zum Beispiel untersuchten Holzinger et al. (2011) in einer Feldstudie den Unterschied zwischen dem Lesen auf einem TFT-LCD und dem Lesen auf Papier im Arbeitskontext einer Klinik. Sie fanden keine Unterschiede hinsichtlich der Lesegeschwindigkeit sowie dem Leseverständnis zwischen den beiden Medien. Unter Laborbedingungen konnten ebenfalls keine Unterschiede zwischen den beiden Medien hinsichtlich Korrekturlesegeschwindigkeit (Gujar, Harrison, & Fishkin, 1998), Leseverständnis sowie Fixationsdauer der Augen beim Lesen und der EEG-Aktivität (Kretzschmar et al., 2013) nachgewiesen werden. Diese Befunde deuten auf eine möglicherweise erreichte Äquivalenz von TFT-LCDs und Papier hin. Es gibt jedoch Untersuchungen, die belegen, dass auch die verbesserte Bildschirmtechnik und die damit einhergehende Optimierung der Bildqualität, die höhere Prävalenz von Augenbeschwerden bei Computernutzern nicht eindämmen konnten (Chu, Rosenfield, & Portello, 2014; Chu, Rosenfield, Portello, Benzoni, & Collier, 2011). Schließlich besteht ebenfalls weiterhin eine starke Präferenz für das Lesen auf Papier im Vergleich zum Lesen auf modernen TFT-LCDs (Holzinger et al., 2011), Tablet Computern und elektronischen Büchern (Kretzschmar et al., 2013). Die Generalisierbarkeit dieser neueren Befunde wird zusätzlich dadurch erschwert, dass wie in den meisten früheren Untersuchungen zum Medienvergleich (eine Ausnahme ist die Untersuchung von Muter et al., 1982) auch in den genannten neueren Untersuchungen ein messwiederholter Ansatz gewählt wurde. Die Problematik bei messwiederholten Studiendesigns besteht darin, dass die Teilnehmer die beiden aufeinanderfolgenden Bedingungen im Hinblick auf ihre Schwierigkeit miteinander vergleichen können. Infolgedessen wird entweder in der vermeintlich schwereren Bedingung die Anstrengungsbereitschaft erhöht oder in der leichteren Bedingung weniger investiert, um ein bestimmtes Leistungsniveau aufrecht zu erhalten. Diese Anpassung kann eine Verschleierung von tatsächlich bestehenden Unterschieden zwischen den zu vergleichenden Bedingungen in Studien mit Messwiederholung verursachen (z.B. Buchner & Baumgartner, 2007). Diese Problematik kann jedoch durch den Einsatz von Gruppendesigns vermieden werden.

Angesichts der deutlichen Fortschritte in der Bildschirmtechnologie und Textdarstellung, der gestiegenen Nutzerexpertise sowie der begrenzten aktuellen Befundlage erscheint ein erneuter, systematischer Vergleich neuer TFT-LCD-Technologie mit Papier notwendig und sinnvoll, um einen möglichen Einfluss auf Leistungsmaße sowie das Wohlbefindens durch die Nutzung von Computerbildschirmen zu untersuchen und eine Aussage über eine mögliche heutige Leistungsäquivalenz der beiden Medien treffen zu können. Zu diesem Zweck wurden vier Experimente durchgeführt, die im Folgenden erläutert werden.

1.1. Experiment 1

In Experiment 1 wurde untersucht, ob sich das Lesen auf modernen TFT-LCDs und auf Papier im Hinblick auf die Lesegeschwindigkeit und die Leistung beim Korrekturlesen, das subjektive Wohlbefinden sowie die Nutzerpräferenz unterscheidet. Hierzu wurden die Lesegeschwindigkeit sowie die Korrekturleseleistung (Anzahl der entdeckten Fehler abzüglich der Anzahl der falschen Alarme) der Teilnehmer beim Lesen an einem modernen TFT-LCD (15.4“, 1680 × 1050 Pixel, 128 ppi Bildschirm mit LED-Hintergrundbeleuchtung eines MacBook Pro) mit dem Lesen auf hochwertigem weißen Papier verglichen. Beide Bedingungen wurden so realisiert, dass sie möglichst äquivalent zueinander waren. So wurden für die Texte sowohl am Bildschirm als auch auf Papier das gleiche Textlayout, die gleiche Schriftart und Schriftgröße, positive Textpolarität sowie die gleiche Methode der Fehlerrückmeldung in der Korrekturleseaufgabe gewählt. Gleichzeitig sollten aber die Eigenschaften einer typischen Lesesituation für beide Lesemedien beibehalten werden. Dazu wurde in der Bildschirmbedingung die bestmögliche Bildschirmauflösung verwendet und durch die Einstellung der höchsten Bildschirmhelligkeit der maximale Helligkeitskontrast gewährleistet. Zudem wurde Subpixel-Rendering aktiviert, um die Textdarstellung zusätzlich zu verfeinern. In der Papierbedingung wurde schwarzer Text in hoher Auflösung auf hochwertiges, weißes Papier gedruckt. Darüber hinaus wurden beide Bedingungen in einem Neigungswinkel präsentiert, der ihrem natürlichen Neigungswinkel in einer typischen Büroumgebung gleichkam. Der Computerbildschirm war entsprechend leicht nach hinten geneigt (75° relativ zur Horizontalen) und das Papier wurde auf einer Buchstütze mit einem Neigungswinkel von 15° präsentiert. Da der bevorzugte Sehabstand beim Lesen auf Papier üblicherweise geringer ist als beim Lesen am Bildschirm (z.B. Gould, Alfaro, Barnes, et al., 1987; Muter & Maurutto, 1991), stand es den Teilnehmern frei, eine möglichst angenehme Sitzposition zu wählen, um so einen optimalen Sehabstand zu gewährleisten.

Die Teilnehmer lasen im ersten Durchgang jeweils sieben Kurzgeschichten in randomisierter Reihenfolge mit einer durchschnittlichen Länge von 870 Wörtern pro Text in der ihnen zugeteilten Bedingung (Bildschirm oder Papier). Die Texte enthielten jeweils 16 Rechtschreibfehler und 14 Grammatikfehler. Die Aufgabe bestand darin, die Texte so genau und schnell wie möglich zu lesen und dabei entdeckte Fehler sowie die dazugehörige Zeilennummer laut zu benennen. Nach einem vorgegebenen Zeitintervall von 3 Minuten pro Text wurden die Teilnehmer aufgefordert, das zuletzt gelesene Wort und die dazu gehörige Zeilennummer zu benennen, und mit dem Lesen des nachfolgenden Textes fortzufahren. Unmittelbar nach dem letzten Text wurde der gewählte Sehabstand der Teilnehmer durch den Versuchsleiter erfasst. Um neben der objektiven Leistung auch einen Eindruck davon zu erhalten, wie die Teilnehmer die Lesesituation beurteilten, wurden die Leistungsmaße der Korrekturleseaufgabe durch subjektive Einschätzungen der Befindlichkeit während der Korrekturleseaufgabe ergänzt. Dazu beantworteten die Teilnehmer zwei Fragebögen zum psychischen und physischen Wohlbefinden im Anschluss an die Leseaufgabe. Die Erfassung direkter Präferenzurteile erforderte einen unmittelbaren Vergleich der beiden Medienbedingungen. Aus diesem Grund lasen die Teilnehmer im Anschluss an den ersten Durchgang weitere sieben Texte in der jeweils anderen Bedingung (Papier oder Bildschirm). Nach dem letzten Text erfolgte erneut die Messung des Sehabstands sowie die Beantwortung der Fragebögen zum Wohlbefinden. Zusätzlich beantworteten die Teilnehmer am Ende des Experiments eine Nachbefragung mit die Fragen zu ihrer Medienpräferenz im Experiment und im Alltag sowie zu ihren allgemeinen Lesegewohnheiten. Es bestand zudem für jeden Teilnehmer die Möglichkeit, Gründe für die jeweiligen Präferenzen im freien Antwortformat anzugeben. Die Ergebnisse der Leistungsmaße sowie der Fragebögen zum subjektiven Wohlbefinden des zweiten Durchgangs wurden auf Grund der genannten möglichen Übertragungsproblematik bei messwiederholten Studiendesigns in der statistischen Auswertung nicht berücksichtigt[1]. Obwohl es dennoch möglich ist, dass sich die Teilnehmer bei

Verwendung eines Gruppendesigns in der vermeintlich leichteren Bedingungen weniger anstrengen als in der schwereren, sind zumindest Übertragungseffekte ausgeschlossen, die auf direkten Vergleichen beruhen. Folglich werden sich Unterschiede zwischen den Bedingungen eher in den Leistungsmaßen widerspiegeln als bei einem Design mit Messwiederholung. Das Gesamtdauer des Experiments betrug etwa 60 Minuten.

Die Ergebnisse zeigten keine signifikanten Unterschiede in der Lesegeschwindigkeit und der Korrekturleseleistung zwischen den beiden Bedingungen (Abbildung 1 und Abbildung 2 in Köpper, Mayr & Buchner, 2015a) und stehen somit im Kontrast zu früheren Befunden, die einen Geschwindigkeitsvorteil von 20-30 % (z.B. Gould & Grischkowsky, 1984) sowie eine bessere Korrekturlesegenauigkeit beim Lesen auf Papier (z.B. Creed et al., 1987) nachweisen konnten. Diese Äquivalenz ist vermutlich auf die verbesserten Bildschirmeigenschaften der modernen TFT-LCDs zurückzuführen. Zum Beispiel war die Bildschirmauflösung in Experiment 1 mit 128 ppi deutlich höher und damit der Punktdichte auf Papier (600 ppi) ähnlicher als die maximal mögliche Auflösung in den frühen Studien (z.B. 62 ppi in Gould & Grischkowsky, 1984). Gestärkt wird diese Vermutung zudem durch eine deskriptiv bessere Lesegeschwindigkeit und Korrekturleseleistung in der Bildschirmbedingung.

Allerdings berichteten die Teilnehmer nach dem Lesen auf dem Bildschirm signifikant mehr Augenbeschwerden (Abbildung 3 in Köpper, Mayr & Buchner, 2015a) und eine stärkere Ermüdung als nach dem Lesen auf Papier. Stärkere Augenbeschwerden in Verbindung mit dem Lesen am Bildschirm wurden einerseits in einigen Untersuchungen mit CRT-Bildschirmen nachgewiesen (z.B. Knave et al., 1985), andererseits scheint diese Problematik auch bei neueren Bildschirmen zu bestehen (Chu et al., 2014; Chu et al., 2011). Dieser Eindruck wird durch die vorliegenden Ergebnisse bestätigt. Eine mögliche Erklärung für diese Problematik ist die erhöhte visuelle Anstrengung, die mit dem Lesen am Bildschirm einhergeht. Zum Beispiel gibt es Belege für eine reduzierte Blinzelhäufigkeit während der Arbeit am Computer im Vergleich zu Ruhebedingungen (Acosta, Gallar, & Belmonte, 1999; Tsubota & Nakamori, 1993) oder computerfreien Aktivitäten, wie Unterhaltungen (Patel, Henderson, Bradley, Galloway, & Hunter, 1991; Schlote, Kadner, & Freudenthaler, 2004). Eine geringere Blinzelhäufigkeit führt zu einer Verschlechterung der Qualität des Tränenfilms und kann juckende und kratzende Augen verursachen (Patel et al., 1991). Zudem gibt es Hinweise auf eine erhöhte Anzahl unvollständiger Lidschläge beim Lesen am Computer im Vergleich zu einer äquivalenten Papierbedingung, was ebenfalls auf eine erhöhte visuelle Anstrengung in der Bildschirmbedingung hindeutet (Chu et al., 2014). Zusätzlich führt der steilere Neigungswinkel des Bildschirms (75°) zu einer Blickrichtung nahe der Horizontalen, während der flachere Neigungswinkel des Papiers (15°) eine stärkere Blickneigung relativ zur Horizontalen bedingt. Dieser Unterschied in der Blickrichtung resultiert in einem größeren Anteil der Augenoberfläche, die der Luft ausgesetzt ist, sowie einer schnelleren Verdunstung der Tränenflüssigkeit in der Bildschirmbedingung (Sotoyama, Jonai, Saito, & Villanueva, 1996; Tsubota & Nakamori, 1993). In Verbindung mit einer reduzierten Blinzelhäufigkeit beim Lesen am Bildschirm lassen sich so die stärkeren Augenbeschwerden in dieser Bedingung erklären.

Eine weitere Folge der unterschiedlichen Neigungswinkel ist vermutlich der Unterschied in den Sehabständen der Teilnehmer. Konsistent zu früheren Befunden (Gould, Alfaro, Barnes, et al., 1987; Muter & Maurutto, 1991) betrug dieser in der Bildschirmbedingung 50 cm und 40 cm in der Papierbedingung.

Die Ergebnisse der Nachbefragung zeigten eine starke Präferenz für die Papierbedingung: Etwa 80 % aller Teilnehmer bevorzugten das Lesen auf Papier, nur 7 % bevorzugten das Lesen am Bildschirm. Dieses Befundmuster bestätigt zum einen die Ergebnisse der frühen Studien mit CRT-Bildschirmen (z.B. Gould, Alfaro, Finn, et al., 1987; Heppner et al., 1985; Oborne & Holton, 1988), aber auch die Ergebnisse der Untersuchungen, die neuere TFT-LCDs verwendeten (Holzinger et al., 2011; Kretzschmar et al., 2013). Die verbesserten Bildschirmeigenschaften der TFT-LCDs, die höhere Vergleichbarkeit der Textdarstellung auf beiden Medien und die dadurch erreichte Äquivalenz in den erhobenen Leistungsmaßen sowie die höhere Vertrautheit der Teilnehmer mit dem Lesen am Bildschirm scheinen folglich die generelle Präferenz für das Lesen auf Papier nicht zu beeinflussen.

1.2. Experiment 2

Die Befunde aus Experiment 1 deuten auf eine Leistungsäquivalenz der Medien in der heutigen Zeit hin, jedoch ist eine Lesezeit von 21 Minuten, wie sie in Experiment 1 realisiert wurde, im Vergleich zu einem gewöhnlichen Arbeitspensum von acht bis neun Stunden Bildschirmarbeit sehr kurz. Es bleibt somit unklar, inwiefern die gezeigte Leistungsäquivalenz auch bei längerer Lesedauer aufrecht erhalten werden kann. Auch die Literatur liefert Hinweise darauf, dass die Lesedauer eine entscheidende Rolle bei der Untersuchung von Leistungsunterschieden zwischen den Medien einnimmt. Zum Beispiel war die Lesedauer pro Medium in Untersuchungen, die einen Unterschied in Leistungsmaßen konnten, deutlich länger als die Lesedauer von 21 Minuten in Experiment 1 (z.B. Gould & Grischkowsky, 1984; Wilkinson & Robinshaw, 1987). In Untersuchungen mit vergleichbar kurzen Lesezeiten konnte hingegen kein Medienunterschied gezeigt werden (z.B. Gujar et al., 1998; Oborne & Holton, 1988). Es ist denkbar, dass Leistungsunterschiede zwischen den Lesemedien über einen bestimmten Zeitraum durch eine Steigerung der Anstrengungsbereitschaft kompensiert werden können. Diese Annahme wird durch Befunde bestärkt, die darauf hindeuten, dass langes Lesen am Bildschirm im Vergleich zum Lesen auf Papier zu einer stärkeren Ermüdung führt (Cushman, 1986; Wilkinson & Robinshaw, 1987).

Die Lesezeit wurde daher in Experiment 2 von 21 auf 63 Minuten verdreifacht. Die Teilnehmer lasen folglich 21 Kurzgeschichten für jeweils 3 Minuten in der ihnen zugeteilten Bedingung (Bildschirm oder Papier) und füllten jeweils vorher und nachher die Fragebögen zu psychischem und physischem Wohlbefinden aus. Am Ende des Experiments erfolgte wie in Experiment 1 eine Nachbefragung, die neben Fragen zum allgemeinen Leseverhalten auch Fragen zur Beurteilung der Lesesituation während des Experiments beinhaltete. Auf die Erfassung direkter Präferenzurteile wurde in Experiment 2 verzichtet.

Das Befundmuster repliziert weitestgehend die Ergebnisse aus Experiment 1. Es zeigten sich keine Unterschiede in der Lesegeschwindigkeit und der Korrekturleseleistung zwischen dem Lesen am Bildschirm und dem Lesen auf Papier (Abbildung 1 und Abbildung 2 in Köpper, Mayr & Buchner, 2015a), jedoch erneut ein stärkerer Anstieg der Augenbeschwerden in der Bildschirmbedingung (Abbildung 3 in Köpper, Mayr & Buchner, 2015a). Zudem wählten die Teilnehmer in der Papierbedingung einen 10 cm kürzeren Sehabstand als die Teilnehmer der Bildschirmbedingung.

Basierend auf den Ergebnissen der Experimente 1 und 2 lässt sich also feststellen, dass die Leistung beim Lesen am Bildschirm mit der Leistung beim Lesen auf Papier vergleichbar und scheinbar nicht auf eine kurzzeitige Kompensation feiner Unterschiede durch eine erhöhte Anstrengungsbereitschaft der Teilnehmer zurückzuführen ist. Dennoch führt das Lesen am Bildschirm weiterhin zu stärken Augenbeschwerden. Es ist offensichtlich, dass weder die Verbesserungen der Bildschirmtechnologie und der Textdarstellung noch die erhöhte Vertrautheit der Nutzer dieses Problem beseitigt haben. Zudem liefert dieser Unterschied eine nachvollziehbare Erklärung für die starke Präferenz für das Lesen auf Papier. Zwei Aspekte wurden von den Teilnehmern von Experiment 1 als Begründung für diese Papierpräferenz besonders häufig genannt: Zum einen beklagten sich die Teilnehmer über eine zu hohe Bildschirmhelligkeit, zum anderen beurteilten sie ihre Leseposition in der Bildschirmbedingung im Vergleich zur Papierbedingung als unangenehmer. Beide Aspekte liefern mögliche Erklärungsansätze für die höhere Prävalenz von Augenbeschwerden in der Bildschirmbedingung und werden daher im Folgenden als Grundlage für die Experimente 3 und 4 erneut aufgegriffen.

1.3. Experiment 3

Hohe Bildschirmhelligkeiten sind mit stärkerer direkter Blendung assoziiert (Isensee, 1982). Diese Blendungswahrnehmungen führen zu unbewussten Kontraktionen der Augenmuskulatur und begünstigen die Entwicklung von Augenbeschwerden bei Tätigkeiten am Bildschirm (Berman, Bullimore, Jacobs, Bailey, & Gandhi, 1994; Thorud et al., 2012). Vor diesem Hintergrund könnten die hohen Bildschirmhelligkeiten verglichen mit der Helligkeit des Papiers in den Experimenten 1 und 2 (Tabelle 1 in Köpper, Mayr & Buchner, 2015a) einen Erklärungsansatz für die stärkeren Augenbeschwerden beim Lesen am Bildschirm darstellen. Um diese Annahme zu testen, wurde in Experiment 3 eine zusätzliche Bildschirmbedingung implementiert, deren Bildschirmhelligkeit an das geringere Helligkeitsniveau der Papierbedingung angeglichen wurde. Falls die Bildschirmhelligkeit in den Experimenten 1 und 2 die Ursache für die stärkeren Augenbeschwerden in der Bildschirmbedingung war, sollten sich die Beschwerden in der neuen Bildschirmbedingung mit reduzierter Bildschirmhelligkeit nicht von den berichteten Beschwerden in der Papierbedingung unterscheiden. Die Bildschirmbedingung mit maximaler Bildschirmhelligkeit wurde beibehalten, um eine Replikation der Befunde aus Experiment 1 und 2 zu ermöglichen. Aufgabe und Durchführung waren analog zu Experiment 2, mit dem Unterschied, dass die Lesezeit auf 45 Minuten reduziert wurde. Dass heißt, die Teilnehmer lasen jeweils 14 Texte in der ihnen zugeteilten Bedingung (Bildschirm, Bildschirm mit reduzierter Bildschirmhelligkeit oder Papier).

Es zeigte sich sowohl in der Bildschirmbedingung mit reduzierter Bildschirmhelligkeit als auch in der regulären Bildschirmbedingung ein signifikant stärkerer Anstieg der Augenbeschwerden im Vergleich zur Papierbedingung (Abbildung 3 in Köpper, Mayr & Buchner, 2015a). Die Teilnehmer schätzten zudem sowohl die Helligkeit als auch die Beeinträchtigung durch Blendung in der helligkeitsreduzierten Bildschirmbedingung höher ein als in der Papierbedingung. Angesichts der Tatsache, dass die absolute Helligkeit der helligkeitsreduzierten Bildschirmbedingung geringfügig niedriger war (ca. 20 cd/m^2) als in der Papierbedingung, ist dieses Befundmuster erstaunlich. Es erscheint jedoch plausibel, dass die inhärenten Lichteigenschaften der beiden Medien die Beurteilungen der Teilnehmer beeinträchtigt haben: Möglicherweise führte die direkte Helligkeit, hervorgerufen durch die Lichtemission des Bildschirms zu einem subjektiv höherem Helligkeitsempfinden als das Licht, das vom Papier in das Auge der Teilnehmer reflektiert wurde.

Auch wenn folglich Einflüsse durch die medienabhängige Lichtemission oder -reflexion nicht vollständig ausgeschlossen werden können, erscheint ausgehend von diesen Ergebnissen die höhere Helligkeit des Bildschirms als mögliche Ursache für die Entstehung und die stärkere Zunahme von Augenbeschwerden beim Lesen am Bildschirm im Vergleich zum Lesen auf Papier unwahrscheinlich. Im Hinblick auf die Lesegeschwindigkeit und Korrekturleseleistung (Abbildung 1 und Abbildung 2 in Köpper, Mayr & Buchner, 2015a) sowie den Sehabstand konnte auch in Experiment 3 das Befundmuster der vorangegangenen Experimente repliziert werden.

1.4. Experiment 4

Neben den Beschwerden über eine zu hohe Bildschirmhelligkeit ergaben die Auswertungen der Ursachen für die Papierpräferenz in Experiment 1, dass die Teilnehmer die Sitzposition während des Experiments in der Papierbedingung als angenehmer empfanden als in der Bildschirmbedingung. Ursächlich für diese unterschiedliche Beurteilung des Sitzkomforts sind möglicherweise die unterschiedlichen Neigungswinkel der beiden Medien (75° in der Bildschirmbedingung gegenüber 15° in der Papierbedingung, relativ zur Horizontalen). Dieser Unterschied liefert zudem eine weitere mögliche Erklärung für die stärkeren Augenbeschwerden in den Bildschirmbedingungen der Experimente 1, 2 und 3: Der steilere Neigungswinkel in der Bildschirmbedingung begünstigt eine Blickrichtung nahe der Horizontalen. Ein geradeaus gerichteter Blick ist mit einer größeren der Luft ausgesetzten Augenoberfläche (Sotoyama et al., 1996), einer entsprechend schnelleren Verdunstung der Tränenflüssigkeit und daraus resultierend, mehr Augenbeschwerden assoziiert (Tsubota & Nakamori, 1993). Im Vergleich dazu ist bei einem flacheren Neigungswinkel, wie in der Papierbedingung, die Blickrichtung stärker geneigt. Dementsprechend ist ein geringerer Anteil der Augenoberfläche exponiert, die Tränenflüssigkeit verdunstet langsamer und das Auftreten von Augenbeschwerden wird verzögert. Folglich sollte ein flacherer Neigungswinkel des Bildschirms zu einer Reduzierung der Augenbeschwerden beitragen und die Medienunterschiede möglicherweise eliminieren.

Um diese Annahme zu testen, wurde in Experiment 4 der Neigungswinkel des Bildschirms an die geringere 15°-Neigung des Papiers angepasst. Da gängige Laptopmodelle über einen zu geringen Aufklappwinkel verfügen, wurde in der Bildschirmbedingung ein Tablet Computer (Apple iPad 2) verwendet, um den flachen Neigungswinkel des Bildschirms zu realisieren. Das Textformat in der Papierbedingung wurde an die geringere Bildschirmgröße des Tablet Computers angepasst. Jeder Teilnehmer las entsprechend 20 Texte für jeweils 2 Minuten bei einer Lesedauer von insgesamt etwa 45 Minuten entweder auf dem Tablet oder auf Papier. Aufgabe und Durchführung waren analog zu den Experimenten 2 und 3.

Tatsächlich unterschied sich die Zunahme der Augenbeschwerden nicht zwischen den beiden Bedingungen (Abbildung 3 in Köpper, Mayr & Buchner, 2015a). Dieses Ergebnis steht im Kontrast zu den Ergebnissen der Experimente 1 - 3 und die Vermutung scheint bestärkt, dass der steilere Neigungswinkel des Bildschirms und die damit einhergehende größere, exponierte Augenoberfläche und die schnellere Verdunstung der Tränenflüssigkeit (Sotoyama et al., 1996; Tsubota & Nakamori, 1993) entscheidende Faktoren bei der Entstehung von Augenbeschwerden beim Lesen am Computerbildschirm sind. Die Auswertung der erhobenen Leistungsmaße ergab darüber hinaus eine signifikant schnellere Lesegeschwindigkeit in der Bildschirmbedingung und es zeigte sich ein deskriptiver Trend hin zu einer besseren Korrekturleseleistung beim Lesen am Bildschirm im Vergleich zum Lesen auf Papier (Abbildung 1 und Abbildung 2 in Köpper, Mayr & Buchner, 2015a). Zudem wählten die Teilnehmer in beiden Bedingungen einen Sehabstand von 40 cm. Dies ist höchstwahrscheinlich auf die identischen Neigungswinkel für Bildschirm und Papier zurückzuführen: Die Teilnehmer passten ihre

Sitzposition in der Bildschirmbedingung an, so dass sich ein Sehabstand, vergleichbar mit dem Sehabstand in der Papierbedingung ergab.

Insgesamt lässt sich auf Grundlage des vierten Experiments festhalten, dass die Leseleistung an TFT-LCDs sogar die Leistung auf Papier übertreffen kann, ohne gleichzeitig stärkere Augenbeschwerden zu verursachen. Voraussetzung dafür ist eine flexible Arbeitsumgebung, die eine papierähnliche Positionierung des Bildschirms ermöglicht.

2. Der Einfluss hoher Bildschirmauflösungen auf Leseverständnis, Korrekturlesegeschwindigkeit und -leistung sowie subjektives Wohlbefinden

Im ersten Teil der Arbeit wurden die Unterschiede zwischen dem Lesen auf modernen TFT-LCDs und dem Lesen auf Papier untersucht. Im Gegensatz zu Studien der 1980er und 1990er Jahre, die vorwiegend CRT-Bildschirme verwendeten, zeigten sich in keinem der hier berichteten Experimente Geschwindigkeitsnachteile oder Leistungseinbußen beim Lesen am TFT-LCD im Vergleich zum Lesen auf Papier. Die Ergebnisse des ersten und vierten Experimentes liefern überdies Hinweise auf eine Überlegenheit des Bildschirms im Hinblick auf die Lesegeschwindigkeit. TFT-LCDs ermöglichen folglich, im Gegensatz zu CRT-Bildschirmen, eine Leseleistung, die der Leistung beim Lesen auf Papier mindestens entspricht. Diese Äquivalenz ist maßgeblich auf die Weiterentwicklung der technischen Bildschirmeigenschaften sowie der digitalen Textdarstellung in den letzten Jahren zurückzuführen, die zu einer deutlichen Verbesserung der Bildschirmqualität und Lesbarkeit von Text am Bildschirm beigetragen haben. Neben anderen Variablen wie der flimmerfreien Präsentation, einer hohen Hintergrundhelligkeit, hohem Helligkeitskontrast sowie verbesserten Möglichkeiten zur Textpräsentation ist die Bildschirmauflösung oder Punktdichte ein Faktor, der für die verbesserte Bildschirmqualität moderner TFT-LCDs verantwortlich ist. Dieser Aspekt und sein möglicher Einfluss auf die objektive Leistungsmaße sowie das subjektive Wohlbefinden stehen im Fokus des zweiten Teils der vorliegenden Arbeit.

In den vergangenen 30 Jahren fand im Bereich der Bildschirmauflösung eine bedeutende Entwicklung statt: Während die ersten CRT-Bildschirme vor 30 Jahren über Bildschirmauflösungen zwischen 60 ppi und 90 ppi verfügten (z.B. Gould, Alfaro, Finn, et al., 1987), besaßen qualitativ hochwertigere CRT-Bildschirme um die Jahrtausendwende bereits Bildschirmauflösungen im Bereich von etwa 100 ppi - 120 ppi (z.B. Alt & Noda, 1998; Ziefle, 1998). Zum heutigen Zeitpunkt sind Bildschirmauflösungen in dieser Höhe der Standard. Die meisten Mobiltelefone und Tabletcomputer sowie einige Laptopbildschirme sind darüber hinaus mit weitaus höher auflösenden Bildschirmen ausgestattet. Beispiele sind die in den Jahren 2010 und 2011 vorgestellten Apple iPads der ersten und zweiten Generation mit jeweils 132 ppi sowie das iPad 3, das über ein sogenanntes "Retinadisplay" mit 264 ppi verfügt. Maßgebliches Ziel dieser Weiterentwicklungen war es, den Informationsgehalt des Bildschirms an die Informationskapazität des menschlichen Auges anzupassen und zum Beispiel die Bildschirmauflösung so zu erhöhen, dass sie dem Auflösungsvermögen des menschlichen Auges entspricht oder es sogar übertrifft (Alt & Noda, 1998). Das "normale" menschliche Auge mit einem Visus von 1.0 kann zwei Bildpunkte mit einem Abstand von 1 Winkelminute (~ 0.017°) unterschieden. Die Größe eines Bildpunktes eines iPad 3 (9.7"; 264 ppi) beträgt etwa 96 µm. Bei einem Sehabstand von 33 cm entspricht diese Größe einem Sehwinkel von 1 Winkelminute. Folglich sind in dieser Entfernung zum Bildschirm die Grenzen des Auflösungsvermögens des menschlichen Auges erreicht und einzelne Bildpunkte können nicht mehr voneinander unterschieden werden.

Während die negativen Effekte einer zu geringen Bildschirmauflösung von CRT-Bildschirmen auf verschiedene Leistungsmaße und auf die Präferenz bekannt sind (z.B. Gould, Alfaro, Finn, et al., 1987; Ziefle, 1998), gibt es bisher nur wenige Untersuchungen zu den möglichen Vorteilen hochauflösender TFT-LCDs mit Bildschirmauflösungen im Bereich des Auflösungsvermögens des menschlichen Auges oder darüber hinaus. Das Befundmuster ist gemischt. Gujar et al. (1998) fanden keine Unterschiede in der Korrekturlesegeschwindigkeit und -genauigkeit sowie dem gewählten Sehabstand der Teilnehmer beim Lesen an einem hochauflösendem Bildschirm (LCD, 282 ppi), einem Bildschirm mit geringer Auflösung (CRT, 85 ppi) sowie auf bedrucktem Papier (300 ppi). Dennoch wurde der subjektive Lesekomfort in der Bedingung mit der geringsten Auflösung (CRT, 85 ppi) im Vergleich zu den anderen Bedingungen am schlechtesten beurteilt. Die Autoren merkten jedoch an, dass Unterschiede in der in Korrekturlesegeschwindigkeit und -genauigkeit möglicherweise durch die geringe Dauer und Komplexität der verwendeten Korrekturleseaufgabe verdeckt wurden. Zudem lässt sich durch die Verwendung verschiedener Bildschirmtechnologien kein eindeutiger Rückschluss auf den Einfluss der Bildschirmauflösung ziehen. Huang, Rau und Liu (2009) wiesen eine geringere Lesegeschwindigkeit bei einer Punktdichte von 125 ppi im Vergleich zu Bildschirmen mit höheren Punktdichten (167 ppi, 200 ppi und 250 ppi) nach, fanden jedoch keinen Einfluss der Punktdichte auf die visuelle Suchleistung. Obwohl in dieser Untersuchung ausschließlich TFT-LCDs verwendet wurden, unterschieden sich die Bildschirme hinsichtlich ihrer Größe und der verwendeten Schriftarten. Ein konfundierender Einfluss dieser Variablen auf die visuelle Leistung kann folglich nicht ausgeschlossen werden. S. L. Wright, Bailey, Tuan und Wacker (1999) fanden keine Unterschiede in der visuellen Suchleistung, der Lesegeschwindigkeit sowie dem Leseverständnis zwischen drei verschiedenen Bildschirmbedingungen (CRT, 102 ppi; TFT-LCD, 83 und 157 ppi). Die Lesbarkeit, operationalisiert durch die Sehschärfe, stieg jedoch mit zunehmender Punktdichte an. Darüber hinaus bevorzugten die Teilnehmer das hochauflösende TFT-LCD gegenüber den anderen Bildschirmbedingungen. Das 83 ppi TFT-LCD wurde im Hinblick auf Präferenz und allgemeinen visuellen Komfort am schlechtesten bewertet. Die Autoren testeten zudem eine weitere Bedingung, in der sie den Sehabstand zum CRT-Bildschirm so vergrößerten, dass das retinale Abbild eines einzelnen Bildpunktes exakt die gleiche Größe hatte wie in der 157 ppi TFT-LCD Bedingung. Zwischen diesen beiden Bedingungen wurden keine Unterschiede in Präferenz und allgemeinem visuellen Komfort nachgewiesen, was darauf hindeutet, dass die Größe des retinalen Abbildes eines Bildpunktes und somit auch die Bildschirmauflösung einen maßgeblichen Einfluss auf die subjektive Beurteilung der Bildschirmqualität sowie den visuellen Komfort, besitzt.

Auf Grundlage des bisherigen Forschungsstands lässt sich ein möglicher Einfluss hoher Bildschirmauflösungen auf die verschiedenen Leistungsmaße zwar vermuten, jedoch ist eine klare Aussage auf Grund der Einschränkungen der genannten Untersuchungen zum jetzigen Zeitpunkt nicht möglich. Aus diesem Grund wurde in den Experimenten 5 und 6 ein TFT-LCD mit einer gängigen Bildschirmauflösung (132 ppi) mit einem hochauflösendem TFT-LCD (264 ppi) verglichen.

2.1. Experiment 5

Das einfache und schnelle Verständnis von digitalem Text ist im Arbeitskontext sowie im Alltag von hoher Relevanz, zum Beispiel beim Lesen von Nachrichten und E-Mails oder beim E-Learning. Entsprechend stellt sich die Frage, inwieweit die erhebliche Verbesserung der Bildschirmauflösung und damit einhergehend die Steigerung der Bild- und Textqualität durch die Zunahme der Bildschärfe zu einer Steigerung des Leseverständnisses führen kann. Zu diesem Zweck wurde in Experiment 5 die Leistung beim Lesen an einem TFT-LCD mit einer gängigen Punktdichte von 132 ppi (9.7", 1024 × 768 pixel, Multi-Touch Display mit LED Hintergrundbeleuchtung und IPS- Technologie eines Apple iPad 2) mit der Leistung an einem hochauflösendem TFT-LCD mit 264 ppi (9.7", 2048 × 1536 pixel, Multi-Touch Retina Display mit LED Hintergrundbeleuchtung und IPS-Technologie eines Apple iPad 3) verglichen[2]. Beide Bedingungen waren mit Ausnahme der Bildschirmauflösung im Hinblick auf entscheidende Bildschirmvariablen, wie Bildschirmgröße, Luminanz, Luminanzkontrast und Kantenglättungsalgorithmen, identisch.

Die Teilnehmer lasen ohne Zeitvorgabe jeweils zwölf Kurzgeschichten in randomisierter Abfolge entweder auf dem 132 ppi oder dem 264 ppi Bildschirm. Im Anschluss an jeden Text folgte ein Forced-Choice Verständnistest mit jeweils fünf inhaltsbezogenen Single-Choice oder Multiple-Choice Fragen. Für jede Frage wurde ein Punktwert berechnet, der in Abhängigkeit von der Anzahl der korrekt oder fälschlich ausgewählten Antwortmöglichkeiten zwischen 0 und 1 variieren konnte, wobei ein Punktwert von 1 eine korrekt beantwortete Frage repräsentierte. In die Auswertung gingen jeweils die Mittelwerte über die Punktwerte der fünf Fragen eines Textes ein.

Auf Grund der sehr hohen Vertrautheit mit dem alltäglichen Lesen auf Papier ist es vorstellbar, dass die Teilnehmer Leistungseinbußen in der schwerer lesbaren Bildschirmbedingung, vermutlich in der Bedingung mit geringerer Bildschirmauflösung, wahrnehmen und im Gegenzug versuchen, ein bestimmtes Leistungsniveau über den Verlauf des Experiments aufrecht zu erhalten. Dieses Verhalten kann die Ergebnisse auf verschiedene Arten verfälschen: Erstens können Verständnisschwierigkeiten beim Lesen eines Textes durch eine entsprechend angepasste Lesegeschwindigkeit ausgeglichen werden (z.B. Cushman, 1986; Dillon, 1992; Mills & Weldon, 1987). Die Interpretation von Unterschieden im Leseverständnis allein ist daher problematisch, da sich Effekte der Bildschirmauflösung auch in unterschiedlichen Lesegeschwindigkeiten der Teilnehmer manifestieren können. Aus diesem Grund wurde in Experiment 5 zusätzlich zum reinen Leseverständnis auch die Lesedauer pro Text erfasst. Zweitens kann eine schlechtere Lesbarkeit eine erhöhte Anstrengungsbereitschaft verursachen, durch die mögliche Effekte der Bildschirmauflösung auf Leseverständnis und Lesedauer verdeckt werden (Buchner & Baumgartner, 2007; Ziefle, 1998). Wie in den vorangegangenen Experimenten wurden daher die Leistungsmaße durch eine Erfassung zusätzlicher Maße des psychologischen und physischen Wohlbefindens vor und nach der Leseaufgabe ergänzt. Schließlich können Lesbarkeitsdefizite durch die Variation des Sehabstandes ausgeglichen werden: Zum einen kann die geringere Punktdichte einen verschwommenen Seheindruck verursachen, dem die Teilnehmer instinktiv durch eine Verringerung ihres Sehabstandes entgegenwirken. Zum anderen verursacht eine Vergrößerung des Sehabstands ein kleineres retinales Abbild einzelner Bildpunkte und könnte somit die subjektiv wahrgenommene Bildschirmqualität sowie den visuellen Komfort steigern (S. L. Wright et al., 1999). Aus diesen Gründen wurde die Sitzposition der Teilnehmer nicht fixiert. Stattdessen wurde der gewählte Sehabstand als zusätzlicher Indikator für die wahrgenommene visuelle Qualität am Ende des Experiments erfasst. Ebenfalls analog zu den Experimenten 2-4 beantworteten die Teilnehmer abschließend die Nachbefragung zu ihren Lesegewohnheiten sowie zur Beurteilung der Lesesituation während des Experiments. Die Gesamtdauer des Experiments variierte in Abhängigkeit von der individuellen Lesegeschwindigkeit der Teilnehmer zwischen 60 und 90 Minuten.

Es zeigten sich keine Unterschiede hinsichtlich des Leseverständnisses zwischen den beiden Bedingungen. Darüber hinaus war auch die mittlere Lesedauer pro Text unabhängig von der Bildschirmauflösung, so dass eine Anpassung der Lesegeschwindigkeit zur Kompensation eines schlechteren Leseverständnisses auszuschließen ist (Abbildung 1 in Köpper, Mayr & Buchner, 2015b). Während also Leistungssteigerungen in früheren Studien bei einer Erhöhung der Punktdichte von CRT-Bildschirmen von 60 ppi auf 80 - 90 ppi gezeigt werden konnten (Gould, Alfaro, Finn, et al., 1987; Ziefle, 1998), deutet das Ausbleiben signifikanter Unterschiede in Leseverständnis und Lesedauer in Experiment 5 darauf hin, dass Punktdichten über 132 ppi keine weitere Verbesserung der Leistung beim Lesen auf TFT-LCDs bewirken. Diese Annahme wurde bereits von Ziefle (1998) geäußert und liefert einen weiteren möglichen Erklärungsansatz für die Befundmuster der Untersuchungen von Gujar et al. (1998), Huang et al. (2009) sowie S. L. Wright et al. (1999). Die Auswertung der Maße zum psychischen und physischem Wohlbefinden sowie die Messungen des Sehabstandes ergaben keine signifikanten Unterschiede zwischen den Bedingungen. Folglich ist eine Alternativerklärung der gefundenen Leistungsäquivalenz durch ein erhöhtes Anstrengungsniveau der Teilnehmer oder eine kompensatorische Anpassung des Sehabstands unwahrscheinlich.

2.2. Experiment 6

Die Ergebnisse von Experiment 5 sprechen dafür, dass die Bildschirmauflösung von TFT-LCDs ab einer bestimmten Höhe keinen Einfluss auf die Leseleistung hat. Es ist jedoch auch denkbar, dass das Leseverständnis als Maß nicht ausreichend sensitiv ist, um mögliche Unterschiede in der Lesbarkeit in Abhängigkeit von der Bildschirmauflösung aufzudecken. Im Vergleich zu Korrekturlesen, visueller Suche oder visuellen Entdeckungsaufgaben ist das Leseverständnis unabhängiger von feinen perzeptuellen Details. Ein hoher Punktwert im Leseverständnis ist dementsprechend nicht nur auf die reine Textverarbeitung, sondern auch auf Wiedererkennen zurückzuführen. Dieser Prozess führt zu einer erhöhten Varianz, die mögliche Effekte der Bildschirmqualität auf das Leseverständnis verdeckt haben könnte. Beim Korrekturlesen ist einerseits eine detailliertere Textverarbeitung erforderlich: Einzelne Wörter und Buchstaben müssen identifiziert werden, um orthographische Fehler zu entdecken. Andererseits ist auch die semantische Verarbeitung des Textes nötig, um Grammatikfehler aufzudecken. Zudem erwiesen sich Korrekturleseaufgaben als geeignet, um Unterschiede in der Leseleistung zwischen Bildschirm und Papier (z.B. Creed et al., 1987; Gould, Alfaro, Barnes, et al., 1987; Gould & Grischkowsky, 1984; Wilkinson & Robinshaw, 1987) sowie zwischen Bildschirmpolaritäten (Buchner & Baumgartner, 2007; Buchner et al., 2009; Creed et al., 1987; Mayr & Buchner, 2010; Piepenbrock, Mayr, & Buchner, 2014a, 2014b; Piepenbrock et al., 2013) aufzudecken. Schließlich ist Korrekturlesen, ebenso wie das Lesen auf Verständnis, eine alltägliche Aufgabe, die eine hohe ökologische Validität garantiert.

Entsprechend war das Ziel von Experiment 6, einen möglichen Effekt der Punktdichte auf die Korrekturlesegeschwindigkeit und -leistung nachzuweisen. Die verwendeten Bildschirme waren identisch zu Experiment 5, Aufbau und Ablauf des Experiments waren identisch zu Experiment 4. Das heißt, die Teilnehmer lasen insgesamt 20 Texte für jeweils 2 Minuten entweder auf dem 132 ppi oder dem 264 ppi Bildschirm. Jeder Text enthielt 20 Fehler (12 Rechtschreibfehler, 8 Grammatikfehler), die verbal und mit Angabe der dazugehörigen Zeilennummer benannt werden sollten. Erfasst wurde hierbei die Lesegeschwindigkeit sowie die Korrekturleseleistung. Nach dem Korrekturlesen wurde der gewählte Sehabstand der Teilnehmer gemessen. Zudem beantworteten die Teilnehmer vor und nach dem Korrekturlesen Fragebögen zu ihrem aktuellen Befinden sowie eine Nachbefragung am Ende des Experiments. Die Gesamtdauer des Experiment betrug etwa 60 Minuten.

Die Lesegeschwindigkeit und die Korrekturleseleistung waren für beide Bildschirmbedingungen gleich hoch (Abbildung 2 in Köpper, Mayr & Buchner, 2015b). Dieses Ergebnis stärkt folglich die Befunde aus Experiment 5 und legt den Schluss nahe, dass die Bildschirmauflösung nur bis zu einer bestimmten Höhe zu einer Verbesserung der Leseleistung führt. Der stärkere Anstieg der Kopfschmerzsymptome sowie der Muskelbeschwerden in der 132 ppi Bedingung über den Verlauf des Experiments deutet jedoch darauf hin, dass die Teilnehmer sich in der 132 ppi Bedingung stärker anstrengen mussten, um ein bestimmtes Geschwindigkeits- und Leistungsniveau aufrecht zu erhalten (Tabelle 3 in Köpper, Mayr & Buchner, 2015b). Der Sehabstand sowie die Beurteilung des Komforts der Leseposition unterschieden sich nicht zwischen den Bedingungen.

3. Allgemeine Diskussion

Seit Einführung der ersten Computer am Arbeitsplatz und in Privathaushalten vor über 30 Jahren wurde das Papier als Lesemedium immer mehr von digitalen Medien verdrängt. Dabei ist die starke Nutzung des Computerbildschirms vornehmlich auf die vielen Vorteile der Digitalisierung zurückzuführen. Diesen Vorteilen standen in den 1980er und 1990er Jahren noch die Nachteile der Computernutzung gegenüber: Die schlechte Bildschirmqualität der frühen CRT-Bildschirme führte zu Einbußen der Lesegeschwindigkeit von 20-30 % (z.B. Gould & Grischkowsky, 1984), Augenbeschwerden sowie muskulären Problemen (z.B. Sheedy, 1992). Die Qualität der Computerbildschirme konnte in den vergangenen Jahren jedoch deutlich gesteigert werden. Veraltete CRT-Bildschirme wurden weitestgehend durch modernere TFT-LCDs ersetzt, die sich durch eine flimmerfreie Bilddarstellung, höhere Bildschirmauflösungen sowie höhere Hintergrundhelligkeiten und Helligkeitskontraste auszeichnen. Weiterentwicklungen der Textdarstellung, wie die Verwendung positiver Polarität, moderner Textverarbeitungsprogramme sowie Kantenglättungsalgorithmen führten zudem zu einer Verbesserung der Lesbarkeit von Text am Bildschirm (z.B. Buchner & Baumgartner, 2007; Sheedy & McCarthy, 1994; Slattery & Rayner, 2010). Ein positiver Einfluss der stetig wachsenden Vertrautheit mit dem Lesen auf digitalen Medien auf die Leseleistung und das Wohlbefinden ist ebenfalls denkbar. Aktuelle Untersuchungen, die diese Neuerungen berücksichtigen, sind jedoch selten und liefern gemischte Befundmuster (Chu et al., 2014; Chu et al., 2011; Gujar et al., 1998; Holzinger et al., 2011; Kretzschmar et al., 2013). Gegenstand des ersten Teils der vorliegenden Arbeit war daher ein erneuter, systematischer Vergleich neuester Bildschirmtechnologien mit konventionellen Papiervorlagen.

In Experiment 1 wurden die Lesegeschwindigkeit, die Korrekturlesegenauigkeit, das subjektive Wohlbefinden sowie die Präferenz beim Lesen auf einem modernen TFT-LCD mit dem Lesen auf Papier verglichen. Der Fokus lag hierbei auf einer möglichst äquivalenten Textdarstellung auf beiden Medien einerseits und einer dem jeweiligen Medium entsprechenden, natürlichen Lesesituation andererseits. Im Gegensatz zu früheren Untersuchungen zeigten sich in Experiment 1 keine Unterschiede in der Korrekturlesegeschwindigkeit und der Korrekturleseleistung zwischen dem Lesen auf einem modernen TFT-LCD und dem Lesen auf Papier. Diese Äquivalenz ist vermutlich auf die verbesserten Bildschirmeigenschaften von TFT-LCDs zurückzuführen. In Konsens mit der vorliegenden aktuelleren Literatur, berichteten die Teilnehmer jedoch nach wie vor stärkere Augenbeschwerden beim Lesen am Bildschirm (Chu et al., 2014; Chu et al., 2011) und äußerten darüber hinaus eine starke Präferenz für das Lesen auf Papier (Holzinger et al., 2011; Kretzschmar et al., 2013).

Um eine mögliche Kompensation kleinerer Unterschiede in den Leistungsmaßen zwischen den Medien durch eine kurzzeitige Erhöhung der Anstrengungsbereitschaft der Teilnehmer zu vermeiden und zudem die ökologische Validität der Aufgabe zu steigern, wurde in Experiment 2 die Lesedauer auf insgesamt etwa eine Stunde verdreifacht. Das Befundmuster aus Experiment 1 konnte bestätigt werden. Es zeigten sich erneut keine Unterschiede in der Korrekturlesegeschwindigkeit und Korrekturleseleistung zwischen TFT-LCD und Papier und die Teilnehmer berichteten stärkere Augenbeschwerden nach dem Lesen auf Papier.

Ausgehend von den ersten beiden Experimenten kann folglich festgehalten werden, dass das Lesen am modernen TFT-LCD keine Einbußen in Lesegeschwindigkeit und -leistung mehr verursacht, jedoch im Vergleich zum Lesen auf Papier weiterhin zu stärker empfundenen Augenbeschwerden führt. Um eine Empfehlung für das Lesen am Bildschirm aussprechen zu können, ist es nötig, die Ursachen für diese Beschwerden zu finden und zu eliminieren. Die Experimente 3 und 4 dienten diesem Zweck.

In Experiment 3 wurde unter Hinzunahme einer weiteren, helligkeitsreduzierten Bildschirmbedingung untersucht, ob die für TFT-LCDs charakteristische hohe Hintergrundhelligkeit einen negativen Einfluss auf die Höhe der Augenbeschwerden hat. Zu hohe Bildschirmhelligkeiten führen zu Blendung (Isensee, 1982) und als Reaktion darauf zu unbewussten Kontraktionen der Augenmuskulatur, die die Entstehung von Augenbeschwerden begünstigen (Berman et al., 1994; Thorud et al., 2012). Es zeigte sich, dass sowohl das Lesen auf dem dunkleren Bildschirm als auch das Lesen auf dem hellen Bildschirm im Vergleich zum Lesen auf Papier zu einer stärkeren Zunahme der Augenbeschwerden führte, so dass ein Einfluss zu hoher Hintergrundhelligkeiten auf die subjektiv berichteten Augenbeschwerden in den ersten beiden Experimenten unwahrscheinlich erscheint. Die Äquivalenz der Medien in den Leistungsmaßen konnte auch in Experiment 3 bestätigt werden.

Eine weitere mögliche Ursache der Augenbeschwerden leitet sich aus den typischen Positionierungen von Bildschirmen und Papier ab. Während Papier, den Empfehlungen entsprechend, in einem Winkel von 15° präsentiert wird, beträgt dieser Winkel bei einem Computerbildschirm typischerweise 75°. Diese Neigungswinkel wurden in den Experimenten 1 - 3 beibehalten, um eine möglichst natürliche Lesesituation zu gewährleisten. Allerdings führt die nahezu senkrechte Ausrichtung des Bildschirms im Vergleich zum flacheren Neigungswinkel des Papiers zu einer horizontaleren Blickrichtung. Dadurch ist ein größerer Anteil der Augenoberfläche exponiert und es verdunstet mehr Tränenflüssigkeit. Diese Faktoren begünstigen maßgeblich die Entstehung von Augenbeschwerden (Sotoyama et al., 1996; Tsubota & Nakamori, 1993). Tatsächlich zeigte sich in Experiment 4 bei einem Neigungswinkel von 15° für beide Medien kein Unterschied in den Augenbeschwerden mehr. Darüber hinaus lasen die Teilnehmer schneller am Bildschirm als auf dem Papier.

Eine Aussage über die tatsächliche Ursache der Augenbeschwerden ist auf Grundlage der hier vorgestellten Experimente vorläufig nicht möglich, da die Angleichung der Neigungswinkel der beiden Medien in Experiment 4 zu weiteren methodischen Veränderungen im Vergleich zu den ersten drei Experimenten führte. So wurde in den Experimenten 1 - 3 der Bildschirm eines MacBook Pro zur Textpräsentation in der Bildschirmbedingung verwendet, aber ein iPad 2 in Experiment 4. Zwar sind diese Bildschirme hinsichtlich der wichtigsten Faktoren der Bildschirmtechnologie sowie der Textdarstellung äquivalent, es gibt jedoch auch Unterschiede, die möglicherweise die Ergebnisse beeinflusst haben könnten. Zu nennen ist erstens der Unterschied in der Bildschirmgröße, da der Laptopbibdschirm mit 15.4“ über eine größere Bildschirmdiagonale verfügt als der Tabletbildschirm mit 9.7“. Die daraus resultierende höhere Helligkeit (etwa 100 lx) am Auge der Teilnehmer beim Lesen am Laptopbildschirm hatte möglicherweise einen Einfluss auf die Höhe der Augenbeschwerden. Vor dem Hintergrund der Ergebnisse des dritten Experimentes, in dem kein Effekt der Bildschirmhelligkeit auf die Höhe der Augenbeschwerden gezeigt werden konnte, erscheint ein solcher Einfluss jedoch vernachlässigbar. Zweitens verfügte das Tablet, jedoch nicht der Laptop-Bildschirm, über IPS (in-plane-switching) Technologie, die zu einer Reduktion der Blickwinkelabhängigkeit von LCDs beiträgt und eine konstantere Bildqualität garantiert, wenn der Bildschirm aus unterschiedlichen Blickrichtungen betrachtet wird. Ein Einfluss dieses Unterschieds auf die Bildqualität kann nicht vollständig ausgeschlossen werden, ist jedoch unwahrscheinlich, da es den Teilnehmern freistand, ihre Lesehaltung selbst zu wählen und somit zu optimieren und die Position der Lesemedien nicht verändert werden durfte. Drittens verfügte der Laptop über Subpixel-Rendering, während das Tablet ausschließlich die Verwendung von Graustufen ermöglichte. Beide Techniken verbessern die Bildqualität (z.B. Sheedy & McCarthy, 1994; Slattery & Rayner, 2010), im direkten Vergleich konnten jedoch keine Leistungsunterschiede nachgewiesen werden (Sheedy, Tai, Subbaram, Gowrisankaran, & Hayes, 2008). Ein Einfluss unterschiedlicher Kantenglättungsalgorithmen scheint also ebenfalls vernachlässigbar.

Eine weitere methodische Veränderung zwischen den Experimenten 1-3 auf der einen Seite und Experiment 4 auf der anderen Seite ergab sich durch die Möglichkeit der Teilnehmer, ihren Sehabstand zum Medium frei zu wählen. Bei büroähnlicher Positionierung der Medien, wie in den Experimenten 1-3, wählten die Teilnehmer konsistent zu früheren Befunden einen Sehabstand von 50 cm in der Bildschirmbedingung und 40 cm in der Papierbedingung (Gould, Alfaro, Barnes, et al., 1987; Muter & Maurutto, 1991), während ein Sehabstand von 40 cm in beiden Bedingungen von Experiment 4 gemessen wurde. Dieser Unterschied resultierte in einer geringfügig größeren retinalen Textgröße in der Papierbedingung (0.29° Sehwinkel) im Vergleich zur Bildschirmbedingung (0.23° Sehwinkel). Unterschiede der Textgröße in diesem Bereich haben jedoch keinen messbaren Einfluss auf die Lesegeschwindigkeit (Legge, 2007), die Beurteilung der Lesbarkeit, Textschärfe und Textschwierigkeit sowie die Präferenz (Bernard, Chaparro, Mills, & Halcomb, 2003). Angesichts der Effektgrößen für den Medienunterschied in den Augenbeschwerden (d = 0.29 - 0.62; d = 0.5 und d = 0.8 sind gemäß Cohen, 1988, als mittlere und große Effekte definiert) in den Experimenten 1 - 3 und den Präferenzurteilen in Experiment 1 (ω = 0.83; ω = 0.5 steht für einen großen Effekt nach Cohen, 1988), ist es folglich unwahrscheinlich, dass die geringfügigen Unterschiede der retinalen Schriftgröße für diese Effekte verantwortlich sind. Auch der umgekehrte Fall, nämlich die gleichen Sehabstände und somit gleich große retinale Textgrößen in beiden Medienbedingungen in Experiment 4 als Erklärung für die Eliminierung der Unterschiede in den Augenbeschwerden, ist dementsprechend fragwürdig. Insgesamt ist ein Einfluss des Sehabstands auf die hier vorgestellten Ergebnisse also unwahrscheinlich.

Zusammengefasst sprechen die Ergebnisse der ersten vier Experimente dafür, dass Lesen an modernen TFT-LCDs so gut ist wie das Lesen auf Papier. Die Problematik erhöhter Augenbeschwerden beim Lesen am Bildschirm scheint bei einer papierähnlichen Positionierung des Bildschirms zu verschwinden. Auch wenn die endgültige Ursache für die Augenbeschwerden noch zu klären ist, ergibt sich daraus die Empfehlung, eine möglichst variable Positionierung von Bildschirmen zu gewährleisten, um der Entstehung von Augenbeschwerden vorzubeugen. Eine Möglichkeit dazu bieten zum Beispiel Tablet-Computer, die einerseits variabel positioniert werden können und andererseits die Vorteile von Papier, wie zum Beispiel die flexiblere Handhabung, mit den Vorteilen von Bildschirmen verbinden. Somit sollten einer uneingeschränkten Nutzung und Verbreitung von Bildschirmtechnologien im Arbeitskontext wie im Privaten keine physischen Beeinträchtigungen mehr entgegenstehen. Angesichts dessen ist eine Weiterentwicklung und Steigerung der Bildschirmqualität sowie eine Anpassung an die Nutzerbedürfnisse besonders wichtig, um eine Optimierung der Leistung beim Lesen am Bildschirm zu erreichen.

Neueste technische Entwicklungen ermöglichen Bildschirmauflösungen, die dem Auflösungsvermögen des menschlichen Auges nahekommen oder es übertreffen. Die Effekte dieser hochauflösenden Bildschirme auf die Leistung sind bisher jedoch nur wenig erforscht (Gujar et al., 1998; Huang et al., 2009; S. L. Wright et al., 1999). Im zweiten Teil der vorliegenden Arbeit sollte daher untersucht werden, in wie weit hohe Bildschirmauflösungen (264 ppi) zu einer Leistungssteigerung im Vergleich zu gängigen Bildschirmauflösungen (132 ppi) beitragen.

In Experiment 5 lag der Fokus zunächst auf der Untersuchung eines möglichen Einflusses der Bildschirmauflösung auf das Leseverständnis. Es zeigten sich keine Unterschiede im Leseverständnis zwischen der 264 ppi und der 132 ppi Bildschirmbedingung. Darüber hinaus konnte die mögliche Kompensation einer Beeinträchtigung des Leseverständnisses durch eine angepasste Lesegeschwindigkeit, eine erhöhte Anstrengung sowie die Variation des Sehabstands weitestgehend ausgeschlossen werden. Es ist jedoch denkbar, dass geringe Lesbarkeitsunterschiede beim Leseverständnis eine untergeordnete Rolle spielen und Aufgaben benötigt werden, die das Erkennen feinerer Textdetails erfordern.

In Experiment 6 wurde daher der Einfluss der Bildschirmauflösung mit einer Korrekturleseaufgabe untersucht. Korrekturlesen erfordert neben der semantischen Verarbeitung des Texts auch die Identifikation einzelner Wörter und Buchstaben und hat sich zudem in zahlreichen Studien als Maß bewährt, um Unterschiede in der Leseleistung aufzudecken (z.B. Buchner & Baumgartner, 2007; Gould &

Grischkowsky, 1984; Piepenbrock et al., 2013; Wilkinson & Robinshaw, 1987). Es zeigten sich keine Unterschiede in der Korrekturlesegeschwindigkeit und der Korrekturleseleistung in Abhängigkeit von der Bildschirmauflösung. Gemeinsam mit Experiment 5 bestätigt dieses Ergebnis die Befundmuster früherer Untersuchungen (Gujar et al., 1998; Huang et al., 2009; S. L. Wright et al., 1999) und spricht dafür, dass die Leseleistung mit zunehmender Punktdichte bis zu einer Höhe von 130-150 ppi ansteigt und dann stagniert. Problematisch ist, dass der Punkt, an dem keine Leistungsverbesserung mehr stattfindet, nicht klar bestimmt ist. Idealerweise sollte dieser durch eine schrittweise Erhöhung der Punktdichte im Bereich von 50 ppi bis etwa 250 ppi ermittelt werden, jedoch sind Bildschirme mit einer Punktdichte unter 100 ppi veraltet und daher nicht mehr verfügbar. Obwohl keine Leistungsunterschiede nachgewiesen wurden, äußerten die Teilnehmer ein höheres Ausmaß an Kopfschmerzsymptomen und Muskelbeschwerden in der 132 ppi Bedingung im Vergleich zur 264 ppi Bedingung, was auf ein erhöhtes Anstrengungsniveau in der 132 ppi Bedingung hinweist. Obwohl Steigerungen der Bildschirmauflösung im Rahmen kürzerer Leseintervalle von bis zu einer Stunde folglich nicht zu einer Leistungssteigerung beitragen, sollten hochauflösende Bildschirme (264 ppi und höher) gegenüber Bildschirmen mit gängigen Bildschirmauflösungen (132 ppi und weniger) bevorzugt werden, um physische Beeinträchtigungen zu vermeiden.

Bei zusammenfassender Betrachtung der vorliegenden Befunde fällt auf, dass sowohl im Vergleich von Bildschirm und Papier als auch im Vergleich der Bildschirmauflösungen (mit Ausnahme der in Experiment 4 dargestellten Befunde) keine Unterschiede in den erhobenen Leistungsmaßen nachgewiesen werden konnten. Die Stichprobengrößen aller hier vorgestellten Experimente waren jedoch ausreichend groß, um mittlere bis große Effekte ($d = 0.6$; $d = 0.5$ und $d = 0.8$ sind definiert als mittlere und große Effekte in den Konventionen von Cohen, 1988) des Mediums bzw. der Bildschirmauflösung mit einer Teststärke von mindestens .95 aufzudecken. Somit ist das Ausbleiben signifikanter Unterschiede in den Leistungsmaßen höchstwahrscheinlich nicht auf einen Typ II-Fehler zurückzuführen, und die Annahme einer heutigen Leistungsäquivalenz wird bestärkt.

Es bleibt ungeklärt, ob sich Leistungsunterschiede in Abhängigkeit vom Medium bzw. der Bildschirmauflösung zeigten, wenn Maße verwendet würden, die den Fokus auf die Identifikation sehr feiner Textdetails legten. Dies lässt sich besonders gut am Beispiel des fünften und sechsten Experiments verdeutlichen. Unabhängig von der Bildschirmauflösung wählten die Teilnehmer beider Experimente einen Sehabstand von 40 cm. In dieser Entfernung beträgt der visuelle Winkel eines Bildpunktes auf dem 132 ppi Bildschirm etwa 0.028°, auf dem 264 ppi Bildschirm hingegen 0.014°. Bei einem Auflösungsvermögen eines „normalen" menschlichen Auges von 0.017° sollten einzelne Bildpunkte auf dem 132 ppi unterscheidbar sein, jedoch nicht auf dem 264 ppi Bildschirm. Obwohl folglich der Auflösungsunterschied zwischen beiden Bedingungen zu Wahrnehmungsunterschieden geführt haben muss, war dies nicht durch Leseverständnis, Lesegeschwindigkeit oder Korrekturleseleistung messbar. S. L. Wright et al. (1999) argumentierten, dass Schriftgrößen nahe der Lesbarkeitsschwelle benötigt würden, um signifikante Unterschiede mit Hilfe der genannten Maße aufzudecken. Allerdings ist es naheliegend, dass Schriftgrößen, die über der Lesbarkeitsschwelle liegen, eine höhere ökologische Validität besitzen als schwellennahe Schriftgrößen. Dies gilt vor allem angesichts der Tatsache, dass normalerweise die Schriftgröße auf Bildschirmen angepasst werden kann.

Zudem ist die Diskrepanz zwischen den Ergebnissen der objektiven und subjektiven Maße zu bemerkenswert. Während nur in Experiment 4 eine signifikant schnellere Lesegeschwindigkeit beim Lesen am Bildschirm nachgewiesen werden konnte, zeigte sich in den meisten hier vorgestellten Experimenten ein Unterschied in den Befindlichkeitsmaßen (Experimente 1 - 3, Experiment 6). Eine mögliche Erklärung für diese Diskrepanz ist eine Erhöhung des Anstrengungsniveaus der Teilnehmer als Reaktion auf eine wahrgenommene geringere Darstellungsqualität, um so einer resultierenden Leistungsbeeinträchtigung entgegenzuwirken. Wenn diese Vermutung zutrifft, sollten Ermüdungserscheinungen beim Lesen am Bildschirm im Vergleich zum Lesen auf Papier sowie bei gering auflösenden Bildschirmen im Vergleich zu hochauflösenden Bildschirmen früher auftreten, so dass sich Leistungsunterschiede zwischen diesen Bedingungen nach einigen Stunden manifestieren sollten. Auch wenn in Experiment 2 keine Effekte einer verlängerten Lesezeit von einer Stunde gezeigt werden konnten, lautet eine mögliche Fragestellung, ob die Leistungsäquivalenz auch nach einem regulären Arbeitstag von acht bis neun Stunden gezeigt werden kann. Dies gilt es in weiteren Untersuchungen zu überprüfen.

Neben einer Generalisierung der hier beschriebenen Befunde auf typischere Arbeitsbedingungen sollte zudem eine Ausweitung der Ergebnisse auf andere Aufgaben sowie Populationen angestrebt werden. Angesichts des voranschreitenden demographischen Wandels und der sukzessiven Erhöhung des Ruhestandsalters erscheint die Frage, ob die hier genannten Schlussfolgerungen auch für ältere Menschen gelten, besonders relevant.

Zusammengefasst sprechen die vorliegenden Ergebnisse dafür, dass Einschränkungen der auf objektive Leistungsmaße sowie des Wohlbefindens beim Lesen an modernen TFT-LCDs nur noch im geringen Ausmaß zu erwarten sind. Es zeigte sich, dass Lesegeschwindigkeit und Korrekturleseleistung beim Lesen von modernen TFT-LCDs so gut ist wie beim Lesen von Papier. Die Problematik erhöhter Augenbeschwerden beim Lesen am Bildschirm scheint bei einer papierähnlichen Positionierung des Bildschirms zu verschwinden, so dass eine flexiblere Handhabung und Positionierung von Bildschirmen empfohlen wird. Maßgeblich verantwortlich für die Äquivalenz von Bildschirm und Papier sind verschiedene Bildschirmfaktoren, wie die Bildschirmauflösung, die in den vergangenen Jahren deutlich verbessert wurden. Bildschirme, die über eine Bildschirmauflösungen verfügen, die das Auflösungsvermögen des menschlichen Auges übertreffen, scheinen zwar nicht zu einer Leistungssteigerung beizutragen, allerdings verringern sie das Risiko von physischen Beschwerden als Folge der Bildschirmarbeit und sollten daher gegenüber Bildschirmen mit herkömmlichen Auflösungen bevorzugt werden.

4. Literaturverzeichnis

Acosta, M. C., Gallar, J., & Belmonte, C. (1999). The influence of eye solutions on blinking and ocular comfort at rest and during work at video display terminals. *Experimental Eye Research, 68*(6), 663-669. doi: 10.1006/exer.1998.0656

Alt, P. M., & Noda, K. (1998). Increasing electronic display information content: An introduction. *IBM Journal of Research and Development, 42*(3.4), 315-320. doi: 10.1147/rd.423.0315

Bergqvist, U., Wolgast, E., Nilsson, B., & Voss, M. (1995). Musculoskeletal disorders among visual display terminal workers: individual, ergonomic, and work organizational factors. *Ergonomics, 38*(4), 763-776. doi: 10.1080/00140139508925148

Berman, S. M., Bullimore, M. A., Jacobs, R. J., Bailey, I. L., & Gandhi, N. (1994). An objective measure of discomfort glare. *Journal of the Illuminating Engineering Society, 23*(2), 40-49. doi: 10.1080/00994480.1994.10748079

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

Buchner, A., & Baumgartner, N. (2007). Text-background polarity affects performance irrespective of ambient illumination and colour contrast. *Ergonomics, 50*(7), 1036-1063. doi: 10.1080/00140130701306413

Buchner, A., Mayr, S., & Brandt, M. (2009). The advantage of positive text-background polarity is due to high display luminance. *Ergonomics, 52*(7), 882-886. doi: 10.1080/00140130802641635

Chen, M.-T., & Lin, C.-C. (2004). Comparison of TFT-LCD and CRT on visual recognition and subjective preference. *International Journal of Industrial Ergonomics, 34*(3), 167-174. doi: 10.1016/j.ergon.2004.02.003

Chu, C., Rosenfield, M., & Portello, J. K. (2014). Blink Patterns: Reading from a Computer Screen versus Hard Copy. *Optometry & Vision Science, 91*(3), 297-302 210.1097/OPX.0000000000000157.

Chu, C., Rosenfield, M., Portello, J. K., Benzoni, J. A., & Collier, J. D. (2011). A comparison of symptoms after viewing text on a computer screen and hardcopy. *Ophthalmic and Physiological Optics, 31*(1), 29-32. doi: 10.1111/j.1475-1313.2010.00802.x

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NY: Erlbaum.

Cole, B. L., Maddocks, J. D., & Sharpe, K. (1996). Effect of VDUs on the eyes: Report of a 6-year epidemiological study. *Optometry and Vision Science, 73*(8), 512-528.

Creed, A., Dennis, I., & Newstead, S. (1987). Proof-reading on VDUs. *Behaviour and Information Technology, 6*(1), 3-13. doi: 10.1080/01449298708901814

Cushman, W. H. (1986). Reading from microfiche, a VDT, and the printed page: Subjective fatigue and performance. *Human Factors, 28*(1), 63-73. doi: 10.1177/001872088602800107

Dillon, A. (1992). Reading from paper versus screens: a critical review of the empirical literature. *Ergonomics, 35*(10), 1297-1326. doi: 10.1080/00140139208967394

Dillon, A., Kleinman, L., Choi, G. O., & Bias, R. (2006). *Visual search and reading tasks using ClearType and regular displays: two experiments*. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Montréal, Québec, Canada.

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. doi: 10.1177/001872088702900303

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. doi: 10.1177/001872088702900501

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. doi: 10.1177/001872088402600308

Gugerty, L., Tyrrell, R. A., Aten, T. R., & Edmonds, K. A. (2004). The effects of subpixel addressing on users' performance and preferences during reading-related tasks. *ACM Trans. Appl. Percept., 1*(2), 81-101. doi: 10.1145/1024083.1024084

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, 42*(6), 527-531.

Harwood, K., & Foley, P. (1987). Temporal Resolution: An Insight into the Video Display Terminal (VDT) "Problem". *Human Factors: The Journal of the Human Factors and Ergonomics Society, 29*(4), 447-452. doi: 10.1177/001872088702900407

Heppner, F. H., Anderson, J. G. T., Farstrup, A. E., & Weiderman, N. H. (1985). Reading performance on a standardized test is better from print than from computer display. *Journal of Reading, 28*(4), 321-325.

Holzinger, A., Baernthaler, M., Pammer, W., Katz, H., Bjelic-Radisic, V., & Ziefle, M. (2011). Investigating paper vs. screen in real-life hospital workflows: Performance contradicts perceived superiority of paper in the user experience. *International Journal of Human-Computer Studies, 69*(9), 563-570. doi: 10.1016/J.Ijhcs.2011.05.002

Huang, D.-L., Rau, P.-L., & Liu, Y. (2009). Effects of font size, display resolution and task type on reading Chinese fonts from mobile devices. *International Journal of Industrial Ergonomics, 39*(1), 81-89. doi: 10.1016/j.ergon.2008.09.004

Hünting, W., Läubli, T., & Grandjean, E. (1981). Postural and visual loads at VDT workplaces I. Constrained postures. *Ergonomics, 24*(12), 917-931. doi: 10.1080/00140138108924914

Isensee, S. (1982). The perception of flicker and glare on computer CRT displays. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 26*(4), 281-284. doi: 10.1177/154193128202600401

Knave, B. G., Wibom, R. I., Voss, M., Hedström, L. D., & Bergqvist, U. O. (1985). Work with video display terminals among office employees. I. Subjective symptoms and discomfort. *Scandinavian Journal of Work, Environment & Health, 11*(6), 457-466. doi: 10.5271/sjweh.2200

Kretzschmar, F., Pleimling, D., Hosemann, J., Füssel, S., Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2013). Subjective impressions do not mirror online reading effort: concurrent EEG-eyetracking evidence from the reading of books and digital media. *PLoS ONE, 8*(2), e56178. doi: 10.1371/journal.pone.0056178

Läubli, T., Hünting, W., & Grandjean, E. (1981). Postural and visual loads at VDT workplaces II. Lighting conditions and visual impairments. *Ergonomics, 24*(12), 933-944. doi: 10.1080/00140138108924915

Legge, G. E. (2007). *Psychophysics of reading in normal and low vision*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Mayr, S., & Buchner, A. (2010). After-effects of TFT-LCD display polarity and display colour on the detection of low-contrast objects. *Ergonomics, 53*(7), 914-925. doi: doi: 10.1080/00140139.2010.484508

Menozzi, M., Lang, F., Näpflin, U., Zeller, C., & Krueger, H. (2001). CRT versus LCD: effects of refresh rate, display technology and background luminance in visual performance. *Displays, 22*(3), 79-85. doi: 10.1016/S0141-9382(01)00054-3

Menozzi, M., Näpflin, U., & Krueger, H. (1999). CRT versus LCD: A pilot study on visual performance and suitability of two display technologies for use in office work. *Displays, 20*(1), 3-10. doi: 10.1016/S0141-9382(98)00051-1

Mills, C. B., & Weldon, L. J. (1987). Reading Text from Computer Screens. *Computing Surveys, 19*(4), 329-358.

Montegut, M. J., Bridgeman, B., & Sykes, J. (1997). High refresh rate and oculomotor adaptation facilitate reading from video displays. *Spatial vision, 10*(4), 305-322. doi: 10.1163/156856897X00230

Muter, P., Latremouille, S. A., Treurniet, W. C., & Beam, P. (1982). Extended reading of continuous text on television screens. *Human Factors, 24*(5), 501-508. doi: 10.1177/001872088202400501

Muter, P., & Maurutto, P. (1991). Reading and skimming from computer screens and books: the paperless office revisited? *Behaviour and Information Technology, 10*(4), 257-266. doi: 10.1080/01449299108924288

Näsänen, R., Karlsson, J., & Ojanpää, H. (2001). Display quality and the speed of visual letter search. *Displays, 22*(4), 107-113. doi: 10.1016/S0141-9382(01)00058-0

Oborne, D. J., & Holton, D. (1988). Reading from screen versus paper: there is no difference. *International Journal of Man-Machine Studies, 28*(1), 1-9. doi: 10.1016/S0020-7373%2888%2980049-X

Ong, C. N., Hoong, B. T., & Phoon, W. O. (1981). Visual and muscular fatigue in operators using visual display terminals. *Journal of human ergology, 10*, 161-171.

Patel, S., Henderson, R., Bradley, L., Galloway, B., & Hunter, L. (1991). Effect of visual display unit use on blink rate and tear stability. *Optometry and Vision Science, 68*(11), 888-892.

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

Piepenbrock, C., Mayr, S., & Buchner, A. (2014b). Smaller pupil size and better proofreading performance with positive than with negative polarity displays. *Ergonomics, 57*(11), 1670-1677. doi: 10.1080/00140139.2014.948496

Piepenbrock, C., Mayr, S., Mund, I., & Buchner, A. (2013). Positive display polarity is advantageous for both younger and older adults. *Ergonomics, 56*(7), 1116-1124. doi: 10.1080/00140139.2013.790485

Rey, P., & Meyer, J. J. (1980). Visual impairments and their objective correlates. In E. Grandjean & E. Vigliani (Eds.), *Ergonomic aspects of visual display terminals* (pp. 77-83). London: Taylor & Francis.

Schlote, T., Kadner, G., & Freudenthaler, N. (2004). Marked reduction and distinct patterns of eye blinking in patients with moderately dry eyes during video display terminal use. *Graefe's Archive for Clinical and Experimental Ophthalmology, 242*(4), 306-312. doi: 10.1007/s00417-003-0845-z

Sheedy, J. E. (1992). Vision problems at video display terminals: a survey of optometrists. *Journal of the American Optometric Association, 63*(10), 687-692.

Sheedy, J. E., & McCarthy, M. (1994). Reading Performance and Visual Comfort with Scale to Gray Compared with Black-and-White Scanned Print. *Displays, 15*(1), 27-30.

Sheedy, J. E., Tai, Y.-C., Subbaram, M., Gowrisankaran, S., & Hayes, J. (2008). ClearType sub-pixel text rendering: Preference, legibility and reading performance. *Displays, 29*(2), 138-151. doi: 10.1016/j.displa.2007.09.016

Shieh, K.-K., & Lin, C.-C. (2000). Effects of screen type, ambient illumination, and color combination on VDT visual performance and subjective preference. *International Journal of Industrial Ergonomics, 26*(5), 527-536. doi: 10.1016/S0169-8141(00)00025-1

Slattery, T. J., & Rayner, K. (2010). The influence of text legibility on eye movements during reading. *Applied Cognitive Psychology, 24*(8), 1129-1148. doi: 10.1002/acp.1623

Smith, M. J., Cohen, B. G. F., Stammerjohn, L. W., & Happ, A. (1981). An investigation of health complaints and job stress in video display operations. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 23*(4), 387-400. doi: 10.1177/001872088102300402

Sotoyama, M., Jonai, H., Saito, S., & Villanueva, M. B. G. (1996). Analysis of ocular surface area for comfortable VDT workstation layout. *Ergonomics, 39*(6), 877-884. doi: 10.1080/00140139608964508

Starr, S. J. (1983). A study of video display terminal workers. *Journal of Occupational and Environmental Medicine, 25*(2), 95-97.

Statistisches Bundesamt. (2012a). *Unternehmen und Arbeitsstätten. Nutzung von Informations- und Kommunikationstechnologien in Unternehmen.* Wiesbaden.

Statistisches Bundesamt. (2012b). *Wirtschaftsrechnungen. Private Haushalte in der Informationsgesellschaft - Nutzung von Informations- und Kommunikationstechnologien.* Wiesbaden.

Thorud, H. M. S., Helland, M., Aarås, A., Kvikstad, T. M., Lindberg, L. G., & Horgen, G. (2012). Eye-related pain induced by visually demanding computer work. *Optometry and Vision Science, 89*(4), 452-464. doi: 10.1097/OPX.0b013e31824c1801

Tsubota, K., & Nakamori, K. (1993). Dry eyes and video display terminals. *New England Journal of Medicine, 328*(8), 584. doi: 10.1056/NEJM199302253280817

Wilkinson, R. T., & Robinshaw, H. M. (1987). Proof-reading: VDU and paper text compared for speed, accuracy and fatigue. *Behaviour and Information Technology, 6*(2), 125-133. doi: 10.1080/01449298708901822

Wright, P., & Lickorish, A. (1983). Proof-reading texts on screen and paper. *Behaviour and Information Technology, 2*(3), 227-235. doi: 10.1080/01449298308914479

Wright, S. L., Bailey, I. L., Tuan, K. M., & Wacker, R. T. (1999). Resolution and legibility: A comparison of TFT-LCDs and CRTs. *Journal of the Society for Information Display, 7*(4), 253-256.

Ziefle, M. (1998). Effects of Display Resolution on Visual Performance. *Human Factors, 40*(4), 554-568. doi: 10.1518/001872098779649355

Ziefle, M. (2001). Aging, visual performance and eyestrain in different screen technologies. *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting*, 262-266. doi: 10.1177/154193120104500315

Endnoten

[1] Eine separate Analyse der Daten dieser Unterstichprobe aus Experiment 1 ist im Anhang von Köpper, Mayr und Buchner (2015a) einzusehen.

[2] Es wurde zusätzlich eine Papierbedingung getestet. Die Ergebnisse dieser Bedingung sind jedoch für die hier zu untersuchende Fragestellung nicht relevant und werden daher nicht berichtet.

Einzelarbeiten

Experimente 1, 2, 3 und 4

Köpper, M., Mayr, S., & Buchner, A. (2015a). Reading from computer screen versus reading from paper: Does it still make a difference? (manuscript submitted for publication).

Experimente 5 und 6

Köpper, M., Mayr, S., & Buchner, A. (2015b). Effects of high pixel density on reading comprehension, proofreading, and subjective well-being (manuscript under review).

Reading from computer screen versus reading from paper: Does it still make a difference?

Maja Köpper, Susanne Mayr, and Axel Buchner

Institut für Experimentelle Psychologie, Heinrich-Heine-Universität, Düsseldorf, Germany

Running head: Reading from screen versus paper

Correspondence concerning this article should be addressed to

Maja Köpper

Institut für Experimentelle Psychologie

Heinrich-Heine-Universität

D-40225 Düsseldorf

Germany

Electronic mail: maja.koepper@hhu.de

(Coauthors' email addresses: susanne.mayr@hhu.de, axel.buchner@hhu.de)

Abstract

Four experiments were conducted to test whether recent developments in display technology would suffice to eliminate the well-known disadvantages in reading from screen as compared with paper. Proofreading speed and performance were equal for a TFT-LCD and a paper display, but there were more symptoms of eyestrain in the screen condition accompanied by a strong preference for paper (Experiment 1). These results were replicated using a longer reading duration (Experiment 2). Additional experiments were conducted to test hypotheses about the reasons for the higher amount of eyestrain associated with reading from screen. Reduced screen luminance did not change the pattern of results (Experiment 3), but positioning both displays in equal inclination angles eliminated the differences in eyestrain symptoms and increased proofreading speed in the screen condition (Experiment 4). A paper-like positioning of TFT-LCDs seems to enable unimpaired reading without evidence of increased physical strain.

Practitioner Summary

Given the developments in screen technology, a re-assessment of the differences in proofreading speed and performance, well-being, and preference between computer screen and paper was conducted. State-of-the-art TFT-LCDs enable unimpaired reading, but a book-like positioning of screens seems necessary to minimise eyestrain symptoms.

Keywords

TFT-LCD, iPad, proofreading, eyestrain, display inclination

Reading from computer screen versus reading from paper: Does it still make a difference?

1. Introduction

Reading text presented on computer screens is an ubiquitous task for most people living in industrialised societies, and it has often replaced reading text printed on paper. Therefore, it seems important to assess how reading from computer screens compares to reading from paper. Early studies from the 1980s have unambiguously shown considerable disadvantages of reading text from screen as compared to paper (Creed, Dennis, & Newstead, 1987; Gould, Alfaro, Barnes, et al., 1987; Gould & Grischkowsky, 1984; Heppner, Anderson, Farstrup, & Weiderman, 1985; Muter, Latremouille, Treurniet, & Beam, 1982; Wilkinson & Robinshaw, 1987; Wright & Lickorish, 1983). For example, Gould and Grischkowsky (1984) found proofreading performance to be 20 - 30% faster on paper than on screen. In a similar study by Wilkinson and Robinshaw (1987), proofreading was both faster and more accurate on paper than on screen.

In addition to the well-documented performance disadvantage, reading from screen was also found to have negative effects on subjective well-being. The main complaints reported by computer users were eye-related symptoms (Cole, Maddocks, & Sharpe, 1996; Knave, Wibom, Voss, Hedström, & Bergqvist, 1985; Läubli, Hünting, & Grandjean, 1981; Ong, Hoong, & Phoon, 1981; Rey & Meyer, 1980; Sheedy, 1992; Smith, Cohen, Stammerjohn, & Happ, 1981) and musculoskeletal pain (Bergqvist, Wolgast, Nilsson, & Voss, 1995; Cole et al., 1996; Hünting, Läubli, & Grandjean, 1981; Ong et al., 1981; Smith et al., 1981; Starr, 1983). For example, Sheedy (1992) conducted a survey on 1307 optometrists who reported that 14.25% of their patients complained about symptoms mainly related to the use of computer screens. The most frequently mentioned symptoms in this context were eye-related symptoms, such as dry eyes, blurred vision or colour vision changes, and musculoskeletal symptoms like neck ache and backache.

However, those empirical findings were obtained the 1980s and 1990s and may no longer be valid for several reasons. First and foremost, all studies mentioned above used cathode ray tube (CRT) display technology. This technology is obsolete and has been mostly replaced by thin film transistor liquid crystal displays (TFT-LCD). Apart from the expedient form factor, the most important advantages of TFT-LCDs are their higher display resolution, their typically higher background luminance (and, hence, higher maximum in text-background luminance contrast), and their flicker-free presentation technique (Menozzi, Lang, Näpflin, Zeller, & Krueger, 2001; Ziefle, 2001). It seems likely that these advancements in screen technology have also lead to improved reading performance and quality (e.g., Gould, Alfaro, Finn, Haupt, & Minuto, 1987). Some recent studies directly compared reading from TFT-LCDs and reading from CRTs and showed a superiority of TFT-LCDs with respect to visual performance (Menozzi et al., 2001; Menozzi, Näpflin, & Krueger, 1999; Näsänen, Karlsson, & Ojanpää, 2001; Shieh & Lin, 2000; Ziefle, 2001) and subjective preference (Chen & Lin, 2004; Shieh & Lin, 2000; Ziefle, 2001). Unfortunately, these studies did not include conditions in which text was read from paper. Therefore, we currently only know that reading from TFT-LCDs is better than reading from CRT displays, but we do not know whether reading from TFT-LCDs is already as good as (or perhaps even better than) reading from paper. Second, very early computers displayed text in negative polarity (i.e., using light letters on a dark background) which is known to result in worse reading performance than text presented in positive polarity (e.g., Buchner & Baumgartner, 2007; e.g., Buchner, Mayr, & Brandt, 2009; e.g., Mayr & Buchner, 2010; e.g., Piepenbrock, Mayr, Mund, & Buchner, 2013). Third, modern operating systems use techniques such as anti-aliasing and subpixel rendering which result in text presentation on computer screens that more closely approximates printed text. Fourth, familiarity may also play a role. For instance, typical participants in studies conducted two to three decades ago presumably had much less experience in reading text from screen as compared to reading text from paper. This may also partially explain the observed performance differences as well as the strong preference for reading on paper that were observed at the time (e.g., Gould, Alfaro, Finn, et al., 1987; e.g., Heppner et al., 1985; e.g., Oborne & Holton, 1988).

Given these significant changes in screen technology and computer user expertise, the comparison between reading from screen and paper needs to be re-assessed. However, recent studies involving modern TFT-LCD technology are very rare and their results do not provide a clear picture. A field study by Holzinger et al. (2011) found that reading speed and text comprehension were no longer impaired for texts presented on modern TFT-LCDs. Similarly, Kretzschmar et al. (2013) found no differences between a modern tablet computer, an electronic book, and paper with regard to reading comprehension. In addition, fixation durations and EEG activity in the theta range, defined individually with reference to a participant's individual alpha frequency (IAF) as IAF-6 Hz to IAF-4 Hz, as online measures of effort in information processing did not differ between the conditions (at least not for younger adults). Nevertheless, there is still a strong preference for reading from paper as compared with reading from a TFT-LCD (Holzinger et al., 2011), from a modern tablet computer or from an electronic book (Kretzschmar et al., 2013). Furthermore, Chu, Rosenfield, Portello, Benzoni, and Collier (2011) compared symptoms of eyestrain after reading from an LCD screen and from paper. Both displays were presented with a viewing distance of 50 cm, an overall luminance of 15 cd/m^2, equivalent text layout, and equal inclination of the paper and the screen display. The authors found a significantly higher symptom score for blurred vision in the screen condition. Most of the other eye-related symptoms (e.g. irritated or burning eyes, dry eyes, eyestrain) were also increased in the screen condition, but these differences were not statistically significant. These findings were confirmed by Chu, Rosenfield, and Portello (2014).

Given this rather limited set of findings, a systematic investigation of the differences between reading from present-day computer screens and paper seemed to be needed. The present series of four experiments pursued exactly this aim. Experiment 1 was designed to test whether there are still differences between computer screen and paper in terms of proofreading speed and performance, subjective well-being, and preference. Experiment 2 was designed to test whether the results of Experiment 1 could be replicated when the reading duration is increased from 25 to 60 minutes. To anticipate, there were no differences in reading performance between screen and paper in both experiments, but symptoms of eyestrain were more pronounced when reading from screen than when reading from paper, and participants exhibited a clear preference for reading from paper. Experiments 3 and 4 were designed to test assumptions about the reasons for the differences in eyestrain symptoms. In Experiment 3, screen luminance was reduced in order to eliminate high screen brightness as a possible cause of the eyestrain symptoms. In Experiment 4, the inclination angles of screen and paper were matched, allowing participants to adopt the same seating posture and gaze direction angle which are known to affect eyestrain symptoms in both conditions.

2. Experiment 1

The purpose of Experiment 1 was to test whether reading from screen still lags behind reading from printed paper or whether modern TFT-LCDs have closed the gap between screen and paper. Specifically, we measured participants' proofreading speed and their proofreading performance for both display conditions. The proofreading task has been frequently used in previous investigations of performance differences between screen and paper (Creed et al., 1987; Gould, Alfaro, Barnes, et al., 1987; Gould & Grischkowsky, 1984; Wilkinson & Robinshaw, 1987). Furthermore, proofreading occurs in many real-life applications and therefore provides measures of high ecological validity.

The screen and paper condition were implemented in a way that guaranteed a maximum degree of equivalence between the two conditions while preserving the core characteristics of a typical reading situation at a modern TFT-LCD and from paper. For both conditions, text layout, font type and size, display polarity, and the method of error annotation in the proofreading task were identical. As is typical for present-day TFT-LCDs, screen presentation was optimised with respect to display resolution, luminance, and luminance contrast. Sub-pixel rendering was activated. This LCD-specific anti-aliasing technique activates each of the three coloured sub-pixels individually to increase the horizontal display resolution and, thereby, to improve the image quality (e.g., Slattery & Rayner, 2010). In the paper condition, black text was printed on white high-quality paper using a 600 dpi laser printer. The inclinations of the display media matched their 'natural' display inclinations in a typical office environment. The computer screen was tilted backwards (75° relative to the horizontal axis) whereas the paper was positioned on a desk stand with a inclination of 15° in order to enable the most comfortable seating posture (Bridger, 1988; Eastman & Kamon, 1976).

Previous research has shown that participants prefer a larger distance when reading from screen than from paper. For example, Gould, Alfaro, Barnes et al. (1987) allowed participants to choose their preferred viewing distance. They found a larger viewing distance for CRT screens (52 cm) than for paper (42 cm). Similarly, Muter and Maurutto (1991) found a very similar difference in viewing distance between screen (58.6 cm) and paper (48 cm). Thus, participants in the present study were asked to choose their preferred seating position and viewing distance without restrictions (such as chin rests) in order to avoid strain due to an uncomfortable posture. The chosen viewing distance was measured by the experimenter at the end of the proofreading task.

Almost all of the studies that compared reading performance between screen and paper used a within-subjects manipulation of display conditions (but see Muter et al., 1982). The problem with within-subjects designs is that participants may compare the conditions with regard to difficulty and either decrease or increase their effort in the seemingly easier or more difficult condition, respectively. As a result, performance differences between the conditions may be obscured in within-subjects designs (e.g., Buchner & Baumgartner, 2007). In order to reduce the problem of performance-effort trade-offs, we used a between-subjects design in which participants experienced only one display condition. Although it is still possible that participants in a between-subjects design invest less or more effort in the supposedly easier or more difficult condition, respectively, at least performance-effort trade-offs based on direct comparisons are impossible. As a consequence, differences in difficulty between conditions will be reflected in performance data more readily in between-subjects than in within-subject designs.

In order to get an idea of how participants experienced the reading situation in their assigned display condition, the performance measurements in the proofreading task were complemented by subjective well-being ratings. We also wanted to know whether participants still prefer reading from paper to reading from screen. To obtain preference judgements based on recent experiences with both media, participants proofread texts from screen and then from paper or vice versa and completed preference questions afterwards. Because this was not the main focus of the study, we collected only preference data from 30 participants in this way. Note that the sole purpose of this data was to provide a basis for the subjective preference judgments. Performance on the second of the two tasks (reading from screen after having read from paper and reading from paper after having read from screen) is no longer independent, but is contaminated by performance on the first task and is thus likely to exhibit the performance-effort trade-offs mentioned above. Therefore, the second reading condition does not provide a usable performance measure which was thus not included in the overall statistical analyses of Experiment 1 (for completeness a brief description of separate statistical analyses of the data of the sub-sample with a second reading task in Experiment 1 is provided in the Appendix).

2.1. Method

2.1.1. Participants

Participants of Experiment 1 were 138 adults, two of whom chose not to follow the proof-reading instructions. Thus, the final sample size was 136. The screen condition comprised 70 participants, 18 of whom were male; they ranged in age from 18 to 33 years (*M* = 23). The paper condition comprised 66 participants, 16 of whom were male; they ranged in age from 19 to 40 years (*M* = 23). The groups differed neither in reading time per day, $\chi^2(4) = 6.25$, $p = .16$, nor in their preference for reading from screen or paper, $\chi^2(4) = 1.78$, $p = .89$.

For all experiments reported here, participants were either paid or received partial course credit, and the inclusion criteria were: Participants had to be native German speakers with normal or corrected-to-normal visual acuity and diagnosed dyslexia. All participants except one were students. Two participants did not fill out any of the questionnaires, one participant did not answer the physical discomfort questionnaire and for another participant the statements about eyestrain were missing. The missing values imply varying degrees of freedom for the different analyses reported.

2.1.2. Material and task

The experiment took place in identical sound-insulated testing chambers without exposure to daylight. The chambers were furnished with a 100 × 100 × 84 cm (width × length × height) table and a height-adjustable office chair. Screen and paper were positioned in their respective standard way of use. The inclination angle of the screen was set at 75° by means of a set square relative to the horizontal axis with the top tilted away from the participant. In the paper condition a desk stand with an inclination of 15° was used. Participants were not allowed to change these preset angles.

The ambient room lighting consisted of two fluorescent lamps and was mainly indirect. Ambient illumination at the participants' approximate eye position (measured with a Gossen Mavolux 5032 B illuminance meter; Gossen Foto- und Lichtmesstechnik GmbH, Nürnberg, Germany; with Class B accuracy according to DIN 5032-7) was 654 lx in the paper condition and 683 lx in the screen condition. Luminance values and luminance contrasts in all four experiments (see Table 1) were determined using a Minolta Colormeter CS-100 (Konica Minolta Co. Ltd, Tokyo, Japan).

please insert Table 1 about here

The proofreading task comprised 14 short stories by various authors. For each participant the order of the stories was random. Each participant proofread seven stories in the condition he or she was assigned to (screen or paper). The remaining seven stories were proofread in the other display condition (paper or screen); this was necessary to assess subjective preferences (see below).

In the screen condition texts were presented by means of HTML pages (Mac OS X, Version 10.6.5; Safari, Version 5.0.3) on the 15.4" (1680 × 1050 pixels, 128 ppi) TFT-LCD widescreen display with LED-backlight of an Apple MacBook Pro which controlled the experiment. The screen area on which the texts were presented comprised 1486 × 1050 pixels. This area does not match the full size of the screen but corresponds to the size of the paper. The remaining screen pixels to the left and right of the text were set to black (RGB = 0,0,0). Subpixel-rendering was activated. In the paper condition each HTML page was printed on a 210 × 297 mm sheet of white high-quality paper (90 g/m^2) using a 600 dpi laser printer.

The mean length of each story was 870 words (*SD* = 9). The stories were presented in 12 pt Helvetica font with an x-height of 2 mm, a cap height of 3 mm, and a line height of 130%. Texts were presented in horizontal page format in two columns with line numbering on both sides. The texts were left-aligned without hyphenation. Each text comprised 60 - 66 lines and 30 - 33 lines in the first and second column, respectively. The maximum line length was 12.6 cm, and the text height was about 11.7 - 12.8 cm. Page layout was basically identical for screen and text displays.

Each short story contained 16 misspellings (4 duplicate letters, 4 missing letters, 4 pairwise letter inversions, 4 incorrect letters) and 14 grammar errors (7 incorrect flexions, 7 incorrect conjugations). The grammar errors forced participants to read for comprehension rather than to scan the text for unusual word shapes (Buchner & Baumgartner, 2007; Creed et al., 1987). Errors were randomly distributed across the text.

Participants were asked to read each text as accurately and fast as possible. Misspellings and grammar errors had to be spoken out loudly together with the line number to disambiguate words that appeared repeatedly in the text. After a reading interval of 3 minutes participants were asked to name the last word which they had read and its corresponding line number before they could start reading the subsequent text. Auditory signals informed participants when to start proof-reading the next text, when half of the proofreading time for the text was over, when the proofreading interval for the text was over, when the last word and the corresponding line number were to be pronounced, and when the page had to be turned.

The two parallel short forms of the multi-dimensional mood state questionnaire (Steyer, Schwenkmezger, Notz, & Eid, 1997) were used to measure three bipolar dimensions of psychological well-being: pleasant versus unpleasant mood, alertness versus fatigue, and calmness versus agitation. The 24 items (12 in each short form) comprise 24 adjectives (e.g. tired, relaxed) as descriptions of the participants' current mood state, which participants rated on a 5-point scale from *not at all* (1) to *very much* (5).

Participants rated physical symptoms of strain using a physical discomfort questionnaire based on the questionnaire developed by Heuer, Hollendiek, Kröger, and Römer (1989). The questionnaire comprised 14 statements about symptoms of eyestrain (7 items), headache (3 items), and musculoskeletal strain (4 items). Participants rated whether they experienced the symptoms on a 7-point scale from *not at all* (1) to *very much* (7).

Finally, participants were asked about their preferred medium in the proofreading task (paper, screen, no preference), about the reasons for the preference, about the amount of time they spent on reading per day (up to 10 minutes, 10 to 30 minutes, 30 to 60 minutes, 1 to 2 hours or more than 2 hours), and whether they generally preferred reading on paper or on screen (exclusively or as often as possible on paper, on paper rather than on screen, no preference, on screen rather than on paper, exclusively or as often as possible on screen; additional free space allowed giving reasons for the preference). There was also free space for any other comments on the experiment.

2.1.3. Procedure

Participants were tested individually. Before starting the experiment they were asked to set the height-adjustable seat to their preferred position. Subsequently, participants were randomly assigned to the screen or the paper condition.

The instructions for the proofreading task were presented on the computer screen, followed by a short practice in which participants proofread a training text for 50 seconds in their assigned display condition. Then, the participants proofread seven texts for 3 minutes each. They started each text at their own discretion. The total reading time was about 25 minutes. Immediately after the end of the last text, participants were asked to keep their current posture until the experimenter had measured the distance between the participants' eyes and the computer or paper display with a tape measure. Next, the participants answered the first parallel short form of the multi-dimensional mood state questionnaire and the physical discomfort questionnaire. Then the experimenter changed the display mode (from screen to paper or vice versa) and the participants started proofreading the remaining seven short stories. Finally, the second parallel short form of the multi-dimensional mood state questionnaire, the physical discomfort questionnaire, and the final questionnaire were administered. The entire experimental procedure took about 60 minutes.

2.1.4. Design

Experiment 1 comprised an one-factorial design with display mode (screen vs. paper) as between-subjects variable. After having finished the proofreading task in the assigned display condition, participants repeated the testing in the other display condition. As mentioned above, this approach was necessary to enable preference ratings of the two different displays. However, the quantitative data of the second testing were not analysed.

The dependent variables were participants' proofreading speed (number of words read) and the proofreading performance (number of errors detected – false alarms), the ratings in the three scales of the multi-dimensional mood state questionnaire (pleasant vs. unpleasant mood, alertness vs. fatigue, calmness vs. agitation), the ratings in the three scales of the physical discomfort questionnaire (eyestrain, headache, musculoskeletal strain), and the chosen viewing distance between participants' eyes and the display.

Assuming a medium to large effect size of $d = 0.60$ for the display mode manipulation (as defined by Cohen, 1988) and desired levels of $\alpha = \beta = .05$, an a priori power analysis revealed that data would have to be collected from $N = 122$ participants (Faul, Erdfelder, Lang, & Buchner, 2007). We were able to collect data from $N = 136$ participants so that the power was even larger than what we had planned for ($1 - \beta = .97$). In all experiments reported here, level of alpha was .05 for all statistical decisions, and all *t*-tests are one-sided.

2.2. Results

2.2.1. Reading speed and proofreading performance

The means of participants' reading speed and proofreading performance are presented in Figure 1 and 2, respectively. Independent groups *t*-tests showed no difference between computer screen and paper in both reading speed, $t(134) = 0.99$, $p = .16$, $d = 0.17$, and proofreading performance, $t(134) = 0.75$, $p = .23$, $d = 0.13$.

please insert Figure 1 about here

please insert Figure 2 about here

2.2.2. Mood state and physical discomfort ratings

The means of participants' evaluations of their current mood state and the statistical results for the screen versus paper comparison are presented in Table 2. Participants reported a significantly stronger degree of fatigue after reading text on the computer screen than after reading text on paper, the remaining dimensions of the mood state questionnaire were not affected by display mode.

please insert Table 2 about here

Table 3 shows the means of participants' physical discomfort ratings and the statistical results for the screen versus paper comparison. Participants reported stronger symptoms of eyestrain after reading text on the computer screen than after reading text on paper (Figure 3). At a descriptive level, there were slightly higher ratings of headache symptoms in the screen than in the paper condition, but this difference just missed the preset level of significance. Participants' evaluations of their musculoskeletal strain were not affected by display mode.

please insert Table 3 about here

please insert Figure 3 about here

2.2.3. Viewing distance

The chosen viewing distance was about 10 cm larger in the screen (M = 50 cm, SE = 1.0 cm) than in the paper condition (M = 40 cm, SE = 0.7 cm), $t(134) = 7.89$, $p < .001$, $d = 1.36$.

2.2.4. Preference judgments

Of the 134 participants who were included in the analysis, 16 perceived the computer screen and paper conditions to be equally pleasant. These 'no preference' answers were not included in the following analysis. 108 participants preferred reading on paper as compared with 10 participants who preferred reading on screen. This imbalance of preferences was statistically significant, $\chi^2(1) = 81.39$, $p < .001$, $\omega = 0.83$, and independent of testing order (paper first–screen second vs. screen first–paper second), $\chi^2(2) = 0.29$, $p = .95$, $\omega = 0.04$. Likewise, participants preferred reading on paper to reading on screen in their daily routine, $\chi^2(1) = 78.03$, $p < .001$, $\omega = 0.21$[1].

2.3. Discussion

In contrast to earlier findings that proofreading on paper may be 20 - 30% faster than proofreading on screen (e.g., Gould & Grischkowsky, 1984), the present Experiment 1 showed no significant differences in proofreading speed and proofreading performance between the screen and paper condition. If anything, there was a non-significant advantage of reading text on a computer screen as compared with reading text on paper. 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 (e.g., 62 ppi in Gould & Grischkowsky, 1984). Furthermore, the TFT-LCD technology provides a flicker-free image, the texts were displayed in positive polarity, and the letters were anti-aliased using subpixel-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.

However, participants' ratings of their physical discomfort revealed significantly stronger symptoms of eyestrain after reading on screen. This finding is in line with earlier studies using CRT technology (Cole et al., 1996; Knave et al., 1985; Läubli et al., 1981; Ong et al., 1981; Rey & Meyer, 1980; Sheedy, 1992; Smith et al., 1981) but also with the results recently reported by Chu et al. (2011) and Chu et al. (2014), who found higher eye-related symptom scores in their LCD conditions. A possible explanation for the stronger ratings of eyestrain in the screen condition is be the increased visual effort associated with reading on screen. For example, there is evidence for a reduced blinking rate during computer work as compared with rest periods (Acosta, Gallar, & Belmonte, 1999; Tsubota & Nakamori, 1993) or computer-free activities such as conversations (Patel, Henderson, Bradley, Galloway, & Hunter, 1991; Schlote, Kadner, & Freudenthaler, 2004). Reduced blinking results in a poorer quality of the tear film and may cause feelings of gritty and itchy eyes (Patel et al., 1991). No differences in blink rate were found between screen and an equivalent paper condition (Chu et al., 2014). However, there was a significant greater number of incomplete blinks in the screen condition which may indicate the increased visual effort invested in this condition. Additionally, given the different inclination angels of screen and paper in the present study (75° vs.

15° relative to the horizontal axis, respectively), the gaze direction was more horizontal when reading from screen than when reading from paper which could have resulted in a larger exposed ocular surface area and an increased rate of tear evaporation in the screen condition (Sotoyama, Jonai, Saito, & Villanueva, 1996; Tsubota & Nakamori, 1993), causing more symptoms of eyestrain. In addition to greater symptoms of eyestrain, participants reported more fatigue after reading from screen than after reading from paper, which is in line with an earlier finding by Cushman (1986) who showed higher ratings of general fatigue after reading on CRT screens as compared to paper.

Participants overwhelmingly preferred the paper condition (80%) over the screen condition (7%). This is consistent with the findings of earlier (e.g., Gould, Alfaro, Finn, et al., 1987; Heppner et al., 1985; Oborne & Holton, 1988) as well as recent studies using modern screen technology (Holzinger et al., 2011; Kretzschmar et al., 2013). This strong preference for paper is surprising given that computer screen quality by now improved to the point at which reading speed and performance no longer differs between screen and paper and given that our participants should have had much more exposure to reading from computer screens. Possible explanations for this will be addressed in the Discussion section of Experiment 2.

Consistent with earlier findings (Gould, Alfaro, Barnes, et al., 1987; Muter & Maurutto, 1991) the preferred viewing distance was 10 cm larger for the screen than for the paper condition. With a 2 mm x-height of the text, the chosen viewing distances of 40 cm and 50 cm resulted in a visual angle of the text of about 0.29° in the paper condition and of 0.23° in the screen condition. However, such a small difference in retinal text size does not seem to have a measurable influence on performance or preference. For example, it is known that reading speed is unaffected within an even broader range of font sizes (0.2° - 2°: see e.g., Legge, 2007). Furthermore, Bernard, Chaparro, Mills and Halcomb (2003) found no differences in subjective ratings of text legibility, sharpness, and difficulty as well as overall preference between a 12 pt Times (~ 0.20°) and a 12 pt Arial (~ 0.25°) font. Given the substantial effect sizes that were found in Experiment 1 for the display mode difference in eyestrain ratings (d = 0.62; d = 0.5 and d = 0.8 are defined as medium and large effects, respectively, in terms of the conventions introduced by Cohen, 1988) and in preference judgments (ω = 0.83; ω = 0.5 is defined as a large effect by Cohen, 1988), it is unlikely that the small differences in visual angle account for these effects.

In Experiment 1, the reading duration was 21 minutes. Compared to previous studies, which found reading speed and performance differences between screen and paper, the reading period was relatively short. For example, in the study by Gould and Grischkowsky (1984) participants read for six work periods of 50 minutes in each of the two display conditions. Two work periods of 50 minutes were scheduled for each display condition in the study by Wilkinson and Robinshaw (1987). In contrast, in the study by Oborne and Holton (1988) who found no difference in reading speed between screen and paper, a reading duration of only 12 minutes per display condition was used. One explanation is that participants were able to compensate for the disadvantages of the screen condition for such a short interval. This would fit with the finding of Wilkinson and Robinshaw (1987) who observed a larger decrement of proofreading accuracy over the course of the experiment in their screen condition and concluded that prolonged reading from screen leads to greater fatigue as compared with reading from paper. Similarly, Cushman (1986) found significantly larger visual fatigue after 60 and 80 minutes of reading on a CRT screen as compared to paper, but no differences after 15, 30, and 45 minutes. Thus, it seems possible that differences in proofreading speed and performance between display conditions take some time to be reflected in observable data because participants may be able to compensate the disadvantages of the computer screen for a while by increased effort. In addition, from the perspective of ecological validity one might argue that a reading period of 21 minutes is too short for a reasonable model task of typical computer work.

Therefore, the purpose of Experiment 2 was to test whether disadvantages of reading on screen would become apparent when the proofreading duration is tripled from 21 to 63 minutes. Due to the increased reading period and the clear results of the preference ratings in Experiment 1, we abstained from testing participants in both display conditions, and, as a consequence, from administering direct preference judgments.

3. Experiment 2

3.1. Method

3.1.1. Participants

Participants were 158 adults. The screen condition comprised 79 participants, 17 of whom were male; they ranged in age from 18 to 39 years (M = 22). The paper condition comprised 79 participants, 21 of whom were male; they ranged in age from 18 to 39 years (M = 23). The groups did not differ in their familiarity with the texts, $\chi^2(1) = 0.18$, $p = .74$, their familiarity with the task, $\chi^2(1) = 0.11$, $p = .85$, their reading time per day, $\chi^2(5) = 5.56$, $p = .35$, or their preference for screen or paper in their daily routine, $\chi^2(4) = 1.79$, $p = .85$.

All participants except two were students. None of them had participated in Experiment 1. Some participants did not answer their questionnaires completely, resulting in four missing values for the ratings of pleasant versus unpleasant mood and one missing value for the ratings of alertness versus fatigue, calmness versus agitation, and the musculoskeletal strain ratings. For two participants the viewing distance measurements were missing.

3.1.2. Material and task

Material and task were identical to those of Experiment 1, with the following exceptions. Ambient illumination at the participants' approximate eye position was 569 lx in the paper condition and 566 lx in the screen condition. The proofreading task comprised 21 new short stories by various authors with a mean length of 881 words (SD = 6). Each participant proofread all 21 short stories in the condition he or she was assigned to (screen or paper). In the final questionnaire participants were asked whether they remembered the texts or the proofreading task from another context[2]. Since participants in Experiment 2 read texts in one of the two conditions only, no direct preference judgements as in Experiment 1 were possible, but participants were asked to assess the reading comfort on a 5-point scale, such as the perceived contrast of the display (too low [1] to too high [5]), the perceived brightness of the display (too low [1] to too high [5]), the quality of their reading position (very uncomfortable [1] to very comfortable [5]), the impairment by glare (not at all [1] to very much [5]), the impairment by reflections (not at all [1] to very much [5].

3.1.3. Procedure

The procedure was identical to that of Experiment 1, with the following exceptions. Participants answered one version of the multi-dimensional mood state questionnaire and the physical discomfort questionnaire before starting the proofreading task (pre), and the other version after proofreading the texts (post). This enabled analyses of the differences between the post- and pre-ratings which provide information about the amount of impairments in mood and physical discomfort as a function of display mode rather than the absolute differences in (post) values in Experiment 1. Every participant read all 21 texts for 3 minutes each. The total reading time, including auditory signals between the texts, was about 68 minutes. In order to enable a temporal orientation, participants received feedback on their progress in the experiment before texts 11, 16, 20, and 21. The entire experimental procedure took about 80 minutes.

3.1.4. Design

The experiment comprised an one-factorial design with display mode (screen vs. paper) as between-subjects variable. The dependent variables were the same as in Experiment 1.

Assuming a medium to large effect size of $d = 0.60$ for the display mode manipulation and desired levels of $\alpha = \beta = .05$, an a priori power analysis revealed that data would have to be collected from $N = 122$ participants. We were able to collect data from $N = 158$ participants so that the power was even larger than what we had planned for ($1 - \beta = .98$).

3.2. Results

3.2.1. Reading speed and proofreading performance

Display mode had no effect on reading speed, $t(156) = 0.09$, $p = .46$, $d = 0.01$, and on proof-reading performance, $t(156) = 0.03$, $p = .49$, $d < 0.01$ (Figures 1 and 2, respectively).

3.2.2. Mood state and physical discomfort ratings

Differences between post- and pre-values for all dimensions of the multi-dimensional mood state questionnaire and the physical discomfort questionnaire were computed. Display mode did not affect changes in any of the three dimensions of the mood state questionnaire (Table 2).

The analyses of the changes in ratings of physical discomfort revealed a stronger increase of eyestrain symptoms after reading on screen as compared with reading on paper. Ratings of head-ache and musculoskeletal strain were not affected by display mode (Table 3).

3.2.3. Viewing Distance

As in Experiment 1, participants chose a larger distance in the screen condition ($M = 54$ cm, $SE = 1$ cm) than in the paper condition ($M = 44$ cm, $SE = 1$ cm), $t(154) = 7.72$, $p < .001$, $d = 1.24$.

3.2.4. Evaluations of reading comfort and preference judgement

The means of participants' evaluations of reading comfort and the statistical results for the screen versus paper comparison are presented in Table 4. Participants rated contrast, impairment by glare, and impairments by reflections to be higher in the screen than in the paper condition. The difference in the ratings of brightness just missed the preset level of significance and there was no difference in the evaluation of the comfort of the reading position. As in Experiment 1, the majority of participants preferred reading from paper in their daily routine, $\chi^2(1) = 51.58$, $p < .001$, $\omega = 0.65$.

3.3. Discussion

The purpose of Experiment 2 was to test whether a prolonged reading duration of about 60 minutes would reveal any disadvantages of the screen condition that might have been obscured by increased effort in the comparatively short 21 minutes reading period of Experiment 1. The pattern of results of Experiment 2 is highly comparable to that of Experiment 1. There were no differences in reading speed and proofreading performance between screen and paper, but a stronger increase in eyestrain ratings in the screen condition. Also similar to Experiment 1, participants' viewing distance was 10 cm shorter in the paper than in the screen condition. The high degree of similarity in the findings despite a much longer reading period in Experiment 2 than in Experiment 1 suggests that the equivalence in reading speed and performance between the screen and paper conditions in Experiment 1 was due to a short-term compensation by increased effort in the screen condition.

Based on the combined pattern of results from Experiments 1 and 2, it seems safe to conclude that performance when reading text on screen has caught up with performance when reading text on paper. However, reading from screens still seems to come at the cost of a stronger experience of eyestrain symptoms. Obviously, neither the improvements in screen technology and software design nor participants' increased computer experience have eliminated this problem, and it thus seems quite reasonable for people still to prefer paper over computer screens for reading.

In order to find possible explanations for the increased eyestrain symptoms in the screen condition and the associated preference for paper, the comments to the preference questions of Experiment 1 were analysed. Unspecific comments such as descriptions of the paper condition as being 'less exhausting' (21 responses) and as being associated with 'less eyestrain' (37 responses) were ignored (these comments essentially confirm the results of the mood state and physical discomfort questionnaires). Two similarly frequent comments were that the display in the screen condition was 'too bright' (22 responses) and that the body posture in the paper condition was 'more comfortable' (21 responses). We decided to test hypotheses derived from these comments in Experiments 3 (screen brightness) and 4 (body posture).

The high screen luminance of the white background in the screen condition of Experiment 1 (of 270 cd/m^2 as opposed to 178 cd/m^2 in the paper condition) might have resulted in more direct glare in the screen condition (Isensee, 1982). Sensation of glare is associated with involuntary contractions of muscles around the eyes and the development of eyestrain symptoms during computer work (Berman, Bullimore, Jacobs, Bailey, & Gandhi, 1994; Thorud et al., 2012). Thus, the higher ratings of eyestrain and fatigue after reading on screen might be a consequence of a high screen luminance. In order to test the screen brightness assumption, an additional screen condition was implemented in Experiment 3. The luminance of this screen condition was reduced relative to the standard screen condition in order to match the lower luminance of the paper condition. If display brightness was the reason why participants reported increased ratings of eyestrain in the screen condition of Experiment 1, the new *screen (luminance reduced)* condition should result in ratings of eyestrain similar to those in the paper condition. The standard (bright) screen condition was also included in order to enable a replication of the results of Experiments 1 and 2. The proofreading results did not differ as a function whether the reading times were 21 and 63 minutes in Experiments 1 and 2, respectively. In order to retain the experimental economy we therefore decided to reduce the reading time to 42 minutes in Experiment 3.

4. Experiment 3

4.1. Method

4.1.1. Participants

Participants were 186 adults. The paper condition comprised 65 participants, 21 of whom were male; they ranged in age from 18 to 38 years (*M* = 24). The screen condition comprised 62 participants, 15 of whom were male; they ranged in age from 18 to 40 years (*M* = 24). The screen (luminance reduced) condition comprised 59 participants, 22 of whom were male; they ranged in age from 18 to 40 years (*M* = 23).

The three groups did not differ with regard to their familiarity with the texts, $\chi^2(2) = 0.41$, *p* =

.81, their familiarity with the task, $\chi^2(2) = 1.28$, $p = .57$, their reading time per day, $\chi^2(10) = 6.37$, $p = .82$, or their preference for screen or paper in their daily routines, $\chi^2(8) = 10.93$, $p = .16$.

All participants except nine were students. Two participants did not answer the multi-dimensional mood state questionnaire and one participant did not answer the physical discomfort questionnaire. For one participant the ratings for alertness versus fatigue were missing. For seven participants the viewing distance measurements were missing.

4.1.2. Material and task

Material and task were identical to those of Experiment 2 with the following exceptions. Ambient illumination at the participants' approximate eye position was 599 lx in the paper condition, 602 lx in the screen condition, and 582 lx in the screen (luminance reduced) condition. The proofreading task comprised the 14 short stories that were used in Experiment 1, but in Experiment 3 all short stories were presented in the same display mode.

4.1.3. Procedure

The procedure was identical to that of Experiment 2, with the following exceptions. Participants read 14 texts for 3 minutes each. The total reading time, including auditory signals between the texts, was about 45 minutes. Auditory feedback on the number of the remaining texts was given before text 8, 13, and 14. The entire experiment took about 60 minutes.

4.1.4. Design

Experiment 3 comprised an one-factorial design with display mode (screen vs. screen [luminance reduced] vs. paper) as between-subjects variable. The dependent variables were the same as in Experiments 1 and 2.

The comparisons between the paper condition and each of the two screen conditions were central to the hypotheses tested. Assuming desired levels of $\alpha = \beta = .05$ and a medium to large effect size of $d = 0.60$ for the difference between the paper and screen condition on the one side and the paper and screen (luminance reduced) condition on the other side, an a priori power analysis revealed that data would have to be collected from $N = 122$ participants for each comparison. We were able to collect data from $N = 127$ participants for the comparison between the screen and paper conditions and $N = 124$ participants for the comparison between the screen (luminance reduced) and paper conditions, implying levels of statistical power of .96 and .95, respectively.

4.2. Results

4.2.1. Reading speed and proofreading performance

Display mode (paper vs. screen) had no effect on reading speed, $t(125) = 0.61$, $p = 0.27$, $d = 0.11$ (Figure 1). Reading speed was descriptively higher in the paper than in the screen (luminance reduced) condition but this difference was not significant, $t(122) = 1.46$, $p = .07$, $d = 0.26$. There was no a difference in proofreading performance between the paper and the screen condition, $t(125) = 0.43$, $p = 0.33$, $d = 0.08$, and between the paper and the screen (luminance reduced) condition, $t(122) = 0.97$, $p = .17$, $d = 0.18$ (Figure 2).

4.2.2. Mood state and physical discomfort ratings

Display mode did not affect changes in any of the three dimensions of the mood state questionnaire (Table 2). As in Experiment 1 and 2, the increase of eyestrain symptoms after reading on screen was stronger than after reading on paper. Likewise, there was a stronger increase of eyestrain symptoms after reading in the screen (luminance reduced) condition than in the paper condition. Headache and musculoskeletal strain ratings were not affected by display mode (Table 3).

4.2.3. Viewing distance

Participants chose a larger distance in the screen condition ($M = 53$ cm, $SE = 1$ cm) and in the screen (luminance reduced) condition ($M = 52$ cm, $SE = 1$ cm) than in the paper condition ($M = 44$ cm, $SE = 1$ cm). The differences between the screen and the paper condition as well as between the screen (luminance reduced) and the paper condition were statistically significant, $t(118) = 5.73$, $p <$

.001, $d = 1.05$ and $t(117) = 5.02$, $p < .001$, $d = 0.93$, respectively.

4.2.4. Evaluations of reading comfort and preference judgement

Participants rated the display brightness as higher in the screen and the screen (luminance reduced) conditions as compared with the paper condition. Similarly, when compared with the paper condition, the impairment by glare was rated to be higher in the screen and in the screen (luminance reduced) condition. There were no differences in participants' evaluations of perceived contrast, comfort of reading position, and impairment by reflections between the paper and screen nor between the paper and screen (luminance reduced) conditions (Table 4). As in the previous experiments, participants had a strong preference for reading on paper in their daily routine, $\chi^2(1) = 81.67$, $p < .001$, $\omega = 0.78$.

4.3. Discussion

The purpose of Experiment 3 was to test whether the particularly large eyestrain symptoms reported after reading on screen in Experiments 1 and 2 were caused by the high screen luminance of the TFT-LCDs. If this were the case, then the screen (luminance reduced) condition should yield ratings of eyestrain that equalled those in the paper condition. This was not the case. The results in the screen (luminance reduced) condition did not differ from those of the standard screen condition. Thus, screen luminance does not seem to be the crucial factor which causes the particularly strong symptoms of eyestrain when reading from screen.

Participants rated the screen brightness and glare in the screen (luminance reduced) condition to be higher than in the paper condition. This is surprising given that the actual luminance of the luminance-reduced screen was even about 20 cd/m^2 lower than the luminance of the paper (Table 1). We have no good explanation for this finding. However, knowing that a screen is a light source whereas paper only reflects light might have biased subjective brightness evaluations.

In the Discussion section of Experiment 2 we reported that participants in Experiment 1 found the body posture 'more comfortable' in the paper than in the screen condition. It seems possible that this is due to the different inclination angles of the screen and paper displays (75° vs. 15° relative to the horizontal axis, respectively). The steeper inclination angle of the computer display presumably resulted in a more horizontal gaze direction angle (evidenced by the larger viewing distance in the screen than in the paper conditions), leading to a larger exposed area of ocular surface, an increased rate of tear evaporation and, hence, more symptoms of eyestrain (Sotoyama et al., 1996; Tsubota & Nakamori, 1993). Thus, the steeper inclination angle of the screen may have caused more symptoms of eyestrain in the screen conditions of Experiment 1, 2, and 3.

The purpose of Experiment 4 was to test whether a screen inclination angle adjusted to that of the paper display would lead to a decrease of eyestrain symptoms in the screen condition, perhaps down to the level of those in the paper condition. A paper-like inclination angle of 15° is impossible for most off-the-shelf laptops due to the limited opening angles of the devices. Therefore, we used a tablet (an Apple iPad 2) which allowed to match the paper display inclination angle of 15°. The screen characteristics of the computer screen used in Experiments 1, 2, and 3 and the iPad 2 screen are highly comparable as both are TFT-LCDs with LED backlight and almost identical display resolutions of 128 and 132 ppi, respectively.

5. Experiment 4

5.1. Method

5.1.1. Participants

Participants were 135 adults, four of whom chose not to follow the proofreading instructions. The final sample comprised 131 participants. The screen condition comprised 69 participants, 15 of whom were male; they ranged in age from 18 to 38 years (M = 24). The paper condition comprised 62 participants, 15 of whom were male; they ranged in age from 19 to 40 years (M = 24).

The groups did not differ with regard to the participants' familiarity with the texts, $\chi^2(1) = 0.17$, $p = .82$, their familiarity with the task, $\chi^2(2) = 4.69$, $p = .09$, their reading time per day, $\chi^2(5) = 3.50$, $p = .65$, and their preference for screen or paper in their daily routine, $\chi^2(3) = 7.19$, $p = .06$. Similarly, there were no differences between the groups regarding the use of an iPad or another tablet, $\chi^2(1) < 0.01$, $p = 1.00$, the use of an iPod touch, iPhone or another smartphone, $\chi^2(1) = 0.06$, $p = .85$, or their experience with those devices, $\chi^2(2) = 3.87$, $p = .14$. All participants except two were students. None of them had participated in Experiments 1, 2, or 3.

One participant did not answer the second short form of the multi-dimensional mood state questionnaire and for another participant the statements about eyestrain of the physical discomfort questionnaire were missing. Five participants did not answer the second page of the final questionnaire and one participant did not answer the final questionnaire at all. For three participants the viewing distance measurements were missing.

5.1.2. Material and task

Material and task were identical to those of Experiment 2 with the following exceptions. Ambient illumination at the participants' approximate eye position was 502 lx in the screen condition and 491 lx in the paper condition. The proofreading task comprised 20 individual short stories by various authors. On average, these texts were 539 (*SD* = 24) words long. In 12 pt Helvetica font the x-height was 2.1 mm and the cap height was 2.8 mm. Each column of text comprised 34 lines. Line length was 8.4 cm and the text height was about 14.1 cm. The layout was exactly the same for both display conditions. Each story contained 12 misspellings (3 duplicate letters, 3 missing letters, 3 pair-wise letter inversions, 3 incorrect letters) and 8 grammar errors (4 incorrect flexions, 4 incorrect conjugations). The reading interval was set to 2 minutes.

In the screen condition texts were typeset in LaTeX and presented as PDF pages in Apple iBooks 1 running under iOS 4.3 on the 9.7" (1024 × 768 pixels, 132 ppi) glossy widescreen multi-touch display with LED backlight and IPS technology of an Apple iPad 2. The texts were anti-aliased. In the paper condition each text was printed on a separate 148 × 210 mm sheet of white paper (90 g/m^2). Texts in the paper condition were a printed versions of the PDF pages.

The final questionnaire contained one additional question in which participants were asked whether they used an iPad or another tablet, whether they used an iPod Touch, and iPhone or another smartphone, and how much experience they had with those devices (none, little, much).

5.1.3. Procedure

The procedure was identical to that of Experiment 2, with the following exceptions. The instructions were presented on a 15.4" MacBook Pro, whereas the proofreading task in the screen condition was presented on the iPad. Participants read 20 texts for 2 minutes each, so that the total reading time, including auditory signals between the texts, was about 45 minutes. The entire experimental procedure took about 60 minutes.

5.1.4. Design

Experiment 4 comprised an one-factorial design with display mode (screen vs. paper) as between-subjects variable. The dependent variables were the same as in Experiments 1-3.

Assuming a medium to large effect size of $d = 0.60$ for the display mode manipulation and desired levels of $\alpha = \beta = .05$, an a priori power analysis revealed that data would have to be collected from $N = 122$ participants. We were able to collect data from $N = 131$ participants so that the power was slightly larger than what we had planned for ($1 - \beta = .96$).

5.2. Results

5.2.1. Reading speed and proofreading performance

Participants read faster on screen than on paper, $t(129) = 2.36$, $p = .01$, $d = 0.42$ (Figure 1). There was a descriptive trend towards better proofreading performance in the screen condition, but this trend was not statistically significant, $t(129) = 1.41$, $p = .08$, $d = 0.25$ (Figure 2).

5.2.2. Mood state and physical discomfort ratings

Participants' increase in agitation was higher after reading on screen than after reading on paper (Table 2). Given that there was no such difference in Experiments 1, 2, and 3, the latter result seems surprising. A more detailed inspection revealed a significant difference between the screen and paper condition for the pre-ratings of the 'calmness versus agitation' scale, $t(129) = 2.68$, $p = .008$ (two-sided), $d = 0.47$, but not for the post-ratings, $t(128) = 0.41$, $p = .68$ (two-sided), $d = 0.07$. Thus, the display mode effect in the post-pre differences is probably due to a priori differences between the groups, but we have no explanation for this difference. The changes in the pleasantness versus unpleasantness and in the calmness versus agitation dimensions were not affected by display mode, and the same was true for the three dimensions of physical discomfort (Table 3).

5.2.3. Viewing distance

The viewing distance in the screen condition ($M = 41$ cm, $SE = 0.8$ cm) did not differ from that of the paper condition ($M = 42$ cm, $SE = 0.9$ cm), $t(126) = 0.25$, $p = .40$, $d = 0.04$.

5.2.4. Preference judgement

As in Experiments 1, 2, and 3, participants expressed a strong preference for reading on paper in their daily routine, $\chi^2(1) = 70.40$, $p < .001$, $\omega = 0.80$.

5.3. Discussion

The purpose of Experiment 4 was to test whether adjusting the inclination angle of the screen condition to that of the paper condition would lead to an elimination of the eyestrain problems. Indeed, there were no longer any differences in eyestrain symptoms between the screen and paper condition (and proofreading was even better in the screen than in the paper condition). This contrasts with the findings of Experiments 1, 2, and 3 which consistently showed a stronger increase of eyestrain symptoms in the screen condition in which the screen inclination angle was 75°, differing from the 15° paper display inclination. Thus, the results support the assumption that display inclination is a crucial factor in the development of eyestrain symptoms, presumably because a gaze direction angle of 75° leads to a larger exposed ocular surface area and an increased rate of tear evaporation (Sotoyama et al., 1996; Tsubota & Nakamori, 1993).

The study by Chu et al. (2011) weakens this conjecture to some extent. The authors taped the text in the paper condition over the screen to ensure equivalent viewing and inclination angles. There were no significant differences between screen and paper for eye-related symptoms such as 'irritated or burning eyes', 'dry eyes', 'eyestrain', 'tired eyes', and 'discomfort in eyes' but descriptively these symptoms were increased in the screen condition. However, the symptom score for 'blurred vision while viewing the text' was significantly higher in the screen than in the paper condition. One could thus speculate that the monitor that was used may not have had the quality of the monitors used here. For instance, monitor properties such as the pixel density are not reported, (although the authors note that the text on the monitor 'appeared to be reasonably clear', p. 32), and a low pixel density could explain the higher prevalence of eye-related symptoms in the screen relative to the paper condition. In addition, Chu et al. argue that because their participants passed both display conditions, they may have been particularly aware of the screen-paper comparison which may have biased their questionnaire responses in the direction of preferences that were based on previous experiences. In the present Experiment 4, we used a between-subjects design which should reduce or even eliminate such biases.

There are some limitations in the comparability of Experiments 1-3 and Experiment 4. The first limitation is that a MacBook Pro was used for text presentation in the screen conditions in Experiments 1-3 but an iPad 2 was used in Experiment 4. The displays of both devices are highly comparable with regard to the following characteristics: Both are TFT-LCDs with LED backlight and almost identical display resolutions (128 ppi of the laptop vs. 132 ppi of the tablet screen) and luminance contrasts (0.96 - 0.98 on the laptop vs. 0.96 on the tablet screen). Further, we presented single pages of text arranged in two columns in all experiments. Font type and size were also equivalent. However, an obvious difference between tablet and laptop screen is the screen size of

9.7″ and 15.4″, respectively. The larger of the two monitors resulted in a higher ambient illumination at the participants' approximate eye position, and this may have affected the eyestrain symptoms. However, Experiment 3 showed no effect of screen brightness on the degree of the perceived eyestrain symptoms which is why it appears unlikely that the reduced amount of light emitted by the tablet in Experiment 4 (as opposed to the laptop in Experiment 3) can account for the equivalence of eyestrain symptoms between tablet screen and paper in Experiment 4. Another difference was that the tablet screen, but not the laptop screen, was equipped with IPS (in-plane-switching) technology that reduces the viewing angle dependency of LCDs and allows for a more constant image quality when the screen is viewed in different orientations. However, participants in the present experiments were not allowed to change the orientations of the reading devices and viewed the screens from optimal positions. Hence, differences in image quality due to IPS technology between laptop and tablet screen are unlikely (but cannot completely be ruled out). Finally, subpixel-rendering was activated on the laptop screen but the tablet only supports grey-scaling (i.e. whole pixels are rendered in shades of grey instead of black or white). Both techniques are known to improve image quality (e.g., Sheedy & McCarthy, 1994; Slattery & Rayner, 2010). Sheedy, Tai, Subbaram, Gowrisankaran and Hayes (2008) did not observe any performance differences when participants worked at a display with activated grey-scaling as compared with a display with activated subpixel-rendering, implying that both techniques are equivalent in their effects. Thus, an influence of the anti-aliasing technique seems to be negligible.

A second limitation is that in Experiments 1-3 the viewing distances were about 40 cm and 50 cm for the paper and the screen conditions, respectively, whereas the preferred viewing distance in Experiment 4 was about 40 cm in both display conditions. As 40 cm was the chosen viewing distance in the paper conditions of Experiments 1-3, the equivalence between screen and paper in Experiment 4 presumably is the result of the adjusted inclination of the tablet screen to a more paper-like inclination of 15°. As a consequence, participants adjusted their seating posture in the screen condition resulting in a shorter viewing distance. As mentioned earlier, the difference in viewing distance and the resulting differences in visual angle of the text is unlikely to account for the higher amount of eyestrain symptoms reported in Experiments 1-3. However, an influence of this difference in visual angle of the text cannot be completely ruled out and additional investigations are necessary, to clarify the underlying reasons for the equivalence in eyestrain in Experiment 4.

Independent of these possible limitations, we may conclude from the results of Experiment 4 that reading speed and proofreading performance with modern TFT-LCD displays may be even better than with paper without the costs of a higher amount of eyestrain provided we allow people to use screens in the same way in which they prefer to use paper. Implications of this finding for application and further research will be mentioned in the general discussion.

6. General Discussion

Early studies from the 1980s were highly consistent in showing disadvantages of reading text from CRT screens as compared with paper. Further, reading from those screens was found to adversely affect subjective well-being, primarily with respect to eye-related symptoms and musculoskeletal pain. State-of-the-art TFT-LCDs outperform obsolete CRT displays in terms of image quality and text presentation capabilities, and reading text from screens should have become much more familiar for many people. Given the rather limited set of recent findings, the goal of the research reported here was to re-assess the comparison between reading from computer screen and paper.

Experiments 1 to 4 consistently showed proofreading on screen to be at least as good as proofreading on paper. It thus seems possible to conclude that reading text on a TFT-LCD screen has caught up with reading text on paper. Note that the sample sizes in the present experiments were relatively large, allowing to detect effects of d = 0.6 (d = 0.5 and d = 0.8 are defined as medium to large effects, respectively, in terms of the conventions introduced by Cohen, 1988) with a statistical power of at least .95 and up to .98. Consequently, the absence of significant differences in visual performance in that order of magnitude is likely not the result of a Type II error. This result stands in contrast to findings of most of the earlier studies and is highly relevant especially with regard to the use of computers for work that requires reading text. Reading-related losses in productivity that were the prize to pay for the many advantages of using computers are unlikely today.

Nevertheless, eyestrain symptoms are more pronounced when reading from screen than when reading from paper. This was shown in Experiments 1, 2, and 3. Experiment 3 showed that the increase in eyestrain symptoms can still be observed in a screen condition in which luminance is reduced relative to that of the paper condition. Thus, the high screen luminance that is typical for state-of-the-art TFT-LCDs does not seem to be the reason for the increase of eyestrain symptoms. In Experiment 4 the differences in eyestrain symptoms between the screen and the paper condition were eliminated by equating the inclination angles of the screen and paper displays. Although additional influences of reading devices cannot completely be ruled out, it seems likely that a paper-like positioning of a screen reduces the occurrence of eyestrain.

The reported equivalence of screen and paper displays for proofreading notwithstanding, participants clearly preferred paper over screens. Such preferences most likely are important factors that determine the perceived quality and appropriateness of displays. In order to get an idea of the perceived advantages of paper in our participants' daily routine, we analysed their reasons for their preference ratings stated in each experiment. The most frequently mentioned reasons for a preference of paper focussed on the reduced experience of eyestrain and exhaustion as well as the more comfortable reading position associated with reading on paper, mobility and flexibility of paper, the possibility to take notes, and the feeling of holding something in one's hands. All of these requirements are met by tablets which may explain their increasing popularity. Thus, in the next years a preference shift from paper towards reading on tablet screen seems likely.

From an applied point of view, this has implications for present-day computer work environments. For example, the introduction of tablet screens at work, especially for long lasting reading tasks, might prevent negative effects of eyestrain on the one hand and combine the perceived advantages of paper with the benefits of screens on the other hand. Consequently, it is the next step to enable a transfer of the present results from the controlled conditions of a laboratory to specific work environments, tasks, and populations. For instance, although Experiment 2 showed no effects of a prolonged reading duration of one hour, one possible question would be whether the performance equivalence extends to many hours of reading in a typical office environment. Another example would be to test whether the conclusions reached here also hold for older persons.

In sum, the present findings support the assumption that reading when using modern TFT-LCD screens is as good as reading when using paper displays. The problem of more complaints about eyestrain when reading from screen than when reading from paper seems to disappear when the screen can be positioned like paper.

7. Footnotes

[1] The options “exclusively and as often as possible on paper” and “on paper rather than on screen” were merged into the classification “preferring paper”. Similarly, the options “exclusively and as often as possible on screen” and “on screen rather than on paper” were merged into the classification “preferring screen”.

[2] Given that (a) the short stories that we used can be found on the internet and (b) that the proofreading task is a frequently used experimental task, participants might have felt familiar with the task or some of the texts. The questions were included to test whether there were any differences in familiarity between the groups.

Table 1

Luminance values (cd/m^2) and luminance contrasts for the paper and screen conditions of Experiment 1, 2, 3, and 4. As luminance measurements in the two chambers differed only marginally—on average, they varied by 5.25% (*SD* = 3.57)—average values are reported.

	Paper			Screen			Screen (luminance reduced)		
	Luminance (cd/m^2)		Contrast[a]	Luminance (cd/m^2)		Contrast[a]	Luminance (cd/m^2)		Contrast[a]
	black	white		black	white		black	white	
Experiment 1	15	178	0.85	3	270	0.98			
Experiment 2	11	157	0.87	4	191	0.96			
Experiment 3	12	161	0.86	3	214	0.97	3	140	0.96
Experiment 4	9	149	0.89	8	349	0.96			

[a] Luminance contrast = $(L_{max} - L_{min})/(L_{max} + L_{min})$

Table 2

Experiment 1: Means, standard errors of the means (lower scores indicate a more negative rating), test statistics for the *t*-tests (*t* and *p*) and effect sizes (*d*) of the three dimensions of the multi-dimensional mood state questionnaire for the first testing. Experiments 2, 3, and 4: Mean post-pre differences, standard errors of the mean differences (negative scores indicate worse mood ratings at the post-measurement), test statistics for the *t*-tests (*t* and *p*) and effect sizes (*d*) of the three dimensions of the multi-dimensional mood state questionnaire.

	Display	*M*	*SE*	*df*	*t*	*p*	*d*
Experiment 1 ($N = 134$, $n_{paper} = 65$, $n_{screen} = 69$)							
pleasant versus unpleasant mood	paper	16.32	0.33	132	0.92	.18	0.16
	screen	15.91	0.30				
alertness versus fatigue	paper	13.54	0.40	**132**	**2.65**	**.005**	**0.46**
	screen	12.09	0.38				
calmness versus agitation	paper	15.97	0.34	132	0.26	.40	0.04
	screen	15.86	0.29				
Experiment 2 ($N = 158$, $n_{paper} = 79$, $n_{screen} = 79$)							
pleasant versus unpleasant mood	paper	-0.29	0.31	152	1.34	.09	0.22
	screen	-0.82	0.25				
alertness versus fatigue	paper	-2.81	0.40	155	0.33	.37	0.05
	screen	-2.63	0.35				
calmness versus agitation	paper	-0.41	0.27	155	0.47	.32	0.08
	screen	-0.45	0.30				
Experiment 3 ($N = 186$, $n_{paper} = 65$, $n_{screen} = 62$, $n_{screen(lr)} = 59$)							
pleasant versus unpleasant mood	paper	-0.29	0.29				
	screen	-0.51	0.22	124[a]	0.59[a]	.28[a]	0.11[a]
	screen (lr)	-0.58	0.28	120[b]	0.71[b]	.24[b]	0.13[b]
alertness versus fatigue	paper	-2.02	0.45				
	screen	-2.77	0.37	125[a]	1.29[a]	.10[a]	0.23[a]
	screen (lr)	-2.32	0.41	120[b]	0.49[b]	.31[b]	0.09[b]
calmness versus agitation	paper	-0.19	0.32				
	screen	0.07	0.28	125[a]	0.59[a]	.28[a]	0.10[a]
	screen (lr)	-0.47	0.28	120[b]	0.67[b]	.25[b]	0.12[b]

Experiment 4 (N = 131, n_{paper} = 62, n_{screen} = 69)							
pleasant versus unpleasant mood	paper	-0.11	0.28	128	0.45	.33	0.08
	tablet	-0.29	0.28				
alertness versus fatigue	paper	-1.15	0.46	128	1.13	.13	0.20
	tablet	-2.19	0.35				
calmness versus agitation	paper	0.75	0.33	**128**	**2.26**	**.01**	**0.40**
	tablet	-0.33	0.34				

Note. screen (lr) = screen (luminance reduced); Significant effects are printed in bold.

[a] paper versus screen. [b] paper versus screen (luminance reduced).

Table 3

Experiment 1: Means, standard errors of the means (larger scores indicate a stronger degree of physical discomfort), test statistics for the *t*-tests (*t* and *p*) and effect sizes (*d*) of the three dimensions of the physical discomfort questionnaire for the first testing. Experiments 2, 3, and 4: Mean post-pre differences, standard errors of the mean differences (positive scores indicate an increase of physical discomfort at the post-measurement), test statistics for the *t*-tests (*t* and *p*) and effect sizes (*d*) of the three dimensions of the physical discomfort questionnaire.

	Display	*M*	*SE*	*df*	*t*	*p*	*d*
Experiment 1 ($N = 134$, $n_{paper} = 65$, $n_{screen} = 69$)							
eyestrain	paper	12.14	0.73	**130**	**3.53**	**.001**	**0.62**
	screen	16.62	1.02				
headache	paper	5.15	0.31	131	1.57	.06	0.27
	screen	6.04	0.45				
musculoskeletal strain	paper	6.52	0.48	131	0.40	.34	0.07
	screen	6.81	0.52				
Experiment 2 ($N = 158$, $n_{paper} = 79$, $n_{screen} = 79$)							
eyestrain	paper	3.49	0.72	**156**	**1.83**	**.035**	**0.29**
	screen	5.38	0.74				
headache	paper	1.48	0.39	156	1.14	.13	0.18
	screen	2.05	0.31				
musculoskeletal strain	paper	1.85	0.44	155	0.24	.40	0.04
	screen	1.71	0.36				
Experiment 3 ($N = 186$, $n_{paper} = 65$, $n_{screen} = 62$, $n_{screen(lr)} = 59$)							
eyestrain	paper	3.25	0.57				
	screen	5.37	0.79	**124**[a]	**2.19**[a]	**.015**[a]	**0.39**[a]
	screen (lr)	5.93	0.75	**121**[b]	**2.88**[b]	**.003**[b]	**0.52**[b]
headache	paper	1.91	0.29				
	screen	1.98	0.34	124[a]	0.18[a]	.43[a]	0.03[a]
	screen (lr)	2.39	0.39	121[b]	1.01[b]	.16[b]	0.18[b]
musculoskeletal strain	paper	1.02	0.24				
	screen	1.37	0.33	124[a]	0.87[a]	.19[a]	0.16[a]
	screen (lr)	0.61	0.28	121[b]	1.12[b]	.13[b]	0.20[b]

Experiment 4 ($N = 131$, $n_{paper} = 62$, $n_{screen} = 69$)							
eyestrain	paper	3.66	0.80	128	0.20	.42	0.04
	tablet	3.45	0.65				
headache	paper	1.86	0.43	129	1.17	.12	0.21
	tablet	1.26	0.29				
musculoskeletal strain	paper	2.08	0.36	129	0.98	.16	0.17
	tablet	1.57	0.38				

Note. screen (lr) = screen (luminance reduced); Significant effects are printed in bold.

[a] paper versus screen. [b] paper versus screen (luminance reduced).

Table 4

Means, standard errors of the means, test statistics for the *t*-tests (*t* and *p*) and effect sizes (*d*) for the evaluations of contrast, brightness, body posture, glare, and reflections for Experiments 2 and 3.

	Display	*M*	*SE*	*df*	*t*	*p*	*d*
Experiment 2 ($N = 158$, $n_{paper} = 79$, $n_{screen} = 79$)							
contrast	paper	3.03	0.05	**153**	**2.78**	**.006**	**0.45**
	screen	3.27	0.07				
brightness	paper	3.22	0.06	152	1.96	.052	0.32
	screen	3.43	0.09				
body posture	paper	3.46	0.08	153	0.48	.63	0.08
	screen	3.51	0.09				
glare	paper	1.65	0.12	**152**	**4.25**	**< .001**	**0.69**
	screen	2.36	0.12				
reflections	paper	1.19	0.06	**150**	**2.46**	**.015**	**0.40**
	screen	1.47	0.10				
Experiment 3 ($N = 186$, $n_{paper} = 65$, $n_{screen} = 62$, $n_{screen(lr)} = 59$)							
contrast	paper	3.05	0.06				
	screen	3.09	0.09	119[a]	0.33[a]	.74[a]	0.03[a]
	screen (lr)	3.05	0.08	117[b]	0.04[b]	.97[b]	0.05[b]
brightness	paper	3.06	0.08				
	screen	3.55	0.10	**121[a]**	**3.81[a]**	**< .001[a]**	**0.69[a]**
	screen (lr)	3.38	0.09	**119[b]**	**2.64[b]**	**.009[b]**	**0.48[b]**
body posture	paper	3.48	0.10				
	screen	3.65	0.08	125[a]	1.30[a]	.20[a]	0.23[a]
	screen (lr)	3.49	0.10	122[b]	0.10[b]	.92[b]	0.02[b]
glare	paper	1.77	0.13				
	screen	2.50	0.14	**122[a]**	**3.91[a]**	**< .001[a]**	**0.70[a]**
	screen (lr)	2.37	0.13	**119[b]**	**3.38[b]**	**.001[b]**	**0.62[b]**
reflections	paper	1.27	0.08				
	screen	1.52	0.11	117[a]	1.75[a]	.08[a]	0.39[a]
	screen (lr)	1.51	0.12	116[b]	1.65[b]	.10[b]	0.44[b]

Note. screen (lr) = screen (luminance reduced); Significant effects are printed in bold.
[a] paper versus screen. [b] paper versus screen (luminance reduced).

8. References

Acosta, M. C., Gallar, J., & Belmonte, C. (1999). The influence of eye solutions on blinking and ocular comfort at rest and during work at video display terminals. *Experimental Eye Research, 68*(6), 663-669. doi: 10.1006/exer.1998.0656

Bergqvist, U., Wolgast, E., Nilsson, B., & Voss, M. (1995). Musculoskeletal disorders among visual display terminal workers: individual, ergonomic, and work organizational factors. *Ergonomics, 38*(4), 763-776. doi: 10.1080/00140139508925148

Berman, S. M., Bullimore, M. A., Jacobs, R. J., Bailey, I. L., & Gandhi, N. (1994). An objective measure of discomfort glare. *Journal of the Illuminating Engineering Society, 23*(2), 40-49. doi: 10.1080/00994480.1994.10748079

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

Bridger, R. S. (1988). Postural adaptations to a sloping chair and work surface. *Human Factors, 30*(2), 237-247. doi: 10.1177/001872088803000210

Buchner, A., & Baumgartner, N. (2007). Text-background polarity affects performance irrespective of ambient illumination and colour contrast. *Ergonomics, 50*(7), 1036-1063. doi: 10.1080/00140130701306413

Buchner, A., Mayr, S., & Brandt, M. (2009). The advantage of positive text-background polarity is due to high display luminance. *Ergonomics, 52*(7), 882-886. doi: 10.1080/00140130802641635

Chen, M.-T., & Lin, C.-C. (2004). Comparison of TFT-LCD and CRT on visual recognition and subjective preference. *International Journal of Industrial Ergonomics, 34*(3), 167-174. doi: 10.1016/j.ergon.2004.02.003

Chu, C., Rosenfield, M., & Portello, J. K. (2014). Blink Patterns: Reading from a Computer Screen versus Hard Copy. *Optometry & Vision Science, 91*(3), 297-302 210.1097/OPX.0000000000000157.

Chu, C., Rosenfield, M., Portello, J. K., Benzoni, J. A., & Collier, J. D. (2011). A comparison of symptoms after viewing text on a computer screen and hardcopy. *Ophthalmic and Physiological Optics, 31*(1), 29-32. doi: 10.1111/j.1475-1313.2010.00802.x

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NY: Erlbaum.

Cole, B. L., Maddocks, J. D., & Sharpe, K. (1996). Effect of VDUs on the eyes: Report of a 6-year epidemiological study. *Optometry and Vision Science, 73*(8), 512-528.

Creed, A., Dennis, I., & Newstead, S. (1987). Proof-reading on VDUs. *Behaviour and Information Technology, 6*(1), 3-13. doi: 10.1080/01449298708901814

Cushman, W. H. (1986). Reading from microfiche, a VDT, and the printed page: Subjective fatigue and performance. *Human Factors, 28*(1), 63-73. doi: 10.1177/001872088602800107

Eastman, M. C., & Kamon, E. (1976). Posture and subjective evaluation at flat and slanted desks. *Human Factors, 18*(1), 15-26. doi: 10.1177/001872087601800102

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191. doi: 10.3758/BF03193146

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. doi: 10.1177/001872088702900303

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. doi: 10.1177/001872088702900501

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. doi: 10.1177/001872088402600308

Heppner, F. H., Anderson, J. G. T., Farstrup, A. E., & Weiderman, N. H. (1985). Reading performance on a standardized test is better from print than from computer display. *Journal of Reading, 28*(4), 321-325.

Heuer, H., Hollendiek, G., Kröger, H., & Römer, T. (1989). Die Ruhelage der Augen und ihr Einfluß auf Beobachtungsabstand und visuelle Ermüdung bei der Bildschirmarbeit. [Resting position of the eyes and its influence on observation distance and visual fatigue during viusal display work]. *Zeitschrift für experimentelle und angewandte Psychologie, 36*(4), 538-566.

Holzinger, A., Baernthaler, M., Pammer, W., Katz, H., Bjelic-Radisic, V., & Ziefle, M. (2011). Investigating paper vs. screen in real-life hospital workflows: Performance contradicts perceived superiority of paper in the user experience. *International Journal of Human-Computer Studies, 69*(9), 563-570. doi: 10.1016/J.Ijhcs.2011.05.002

Hünting, W., Läubli, T., & Grandjean, E. (1981). Postural and visual loads at VDT workplaces I. Constrained postures. *Ergonomics, 24*(12), 917-931. doi: 10.1080/00140138108924914

Isensee, S. (1982). The perception of flicker and glare on computer CRT displays. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 26*(4), 281-284. doi: 10.1177/154193128202600401

Knave, B. G., Wibom, R. I., Voss, M., Hedström, L. D., & Bergqvist, U. O. (1985). Work with video display terminals among office employees. I. Subjective symptoms and discomfort.

Scandinavian Journal of Work, Environment & Health, 11(6), 457-466. doi: 10.5271/sjweh.2200

Kretzschmar, F., Pleimling, D., Hosemann, J., Füssel, S., Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2013). Subjective impressions do not mirror online reading effort: concurrent EEG-eyetracking evidence from the reading of books and digital media. *PLoS ONE, 8*(2), e56178. doi: 10.1371/journal.pone.0056178

Läubli, T., Hünting, W., & Grandjean, E. (1981). Postural and visual loads at VDT workplaces II. Lighting conditions and visual impairments. *Ergonomics, 24*(12), 933-944. doi: 10.1080/00140138108924915

Legge, G. E. (2007). *Psychophysics of reading in normal and low vision*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Mayr, S., & Buchner, A. (2010). After-effects of TFT-LCD display polarity and display colour on the detection of low-contrast objects. *Ergonomics, 53*(7), 914-925. doi: doi: 10.1080/00140139.2010.484508

Menozzi, M., Lang, F., Näpflin, U., Zeller, C., & Krueger, H. (2001). CRT versus LCD: effects of refresh rate, display technology and background luminance in visual performance. *Displays, 22*(3), 79-85. doi: 10.1016/S0141-9382(01)00054-3

Menozzi, M., Näpflin, U., & Krueger, H. (1999). CRT versus LCD: A pilot study on visual performance and suitability of two display technologies for use in office work. *Displays, 20*(1), 3-10. doi: 10.1016/S0141-9382(98)00051-1

Muter, P., Latremouille, S. A., Treurniet, W. C., & Beam, P. (1982). Extended reading of continuous text on television screens. *Human Factors, 24*(5), 501-508. doi: 10.1177/001872088202400501

Muter, P., & Maurutto, P. (1991). Reading and skimming from computer screens and books: the paperless office revisited? *Behaviour and Information Technology, 10*(4), 257-266. doi: 10.1080/01449299108924288

Näsänen, R., Karlsson, J., & Ojanpää, H. (2001). Display quality and the speed of visual letter search. *Displays, 22*(4), 107-113. doi: 10.1016/S0141-9382(01)00058-0

Oborne, D. J., & Holton, D. (1988). Reading from screen versus paper: there is no difference. *International Journal of Man-Machine Studies, 28*(1), 1-9. doi: 10.1016/S0020-7373%2888%2980049-X

Ong, C. N., Hoong, B. T., & Phoon, W. O. (1981). Visual and muscular fatigue in operators using visual display terminals. *Journal of human ergology, 10*, 161-171.

Patel, S., Henderson, R., Bradley, L., Galloway, B., & Hunter, L. (1991). Effect of visual display unit use on blink rate and tear stability. *Optometry and Vision Science, 68*(11), 888-892.

Piepenbrock, C., Mayr, S., Mund, I., & Buchner, A. (2013). Positive display polarity is advantageous for both younger and older adults. *Ergonomics, 56*(7), 1116-1124. doi: 10.1080/00140139.2013.790485

Rey, P., & Meyer, J. J. (1980). Visual impairments and their objective correlates. In E. Grandjean & E. Vigliani (Eds.), *Ergonomic aspects of visual display terminals* (pp. 77-83). London: Taylor & Francis.

Schlote, T., Kadner, G., & Freudenthaler, N. (2004). Marked reduction and distinct patterns of eye blinking in patients with moderately dry eyes during video display terminal use. *Graefe's Archive for Clinical and Experimental Ophthalmology, 242*(4), 306-312. doi: 10.1007/s00417-003-0845-z

Sheedy, J. E. (1992). Vision problems at video display terminals: a survey of optometrists. *Journal of the American Optometric Association, 63*(10), 687-692.

Sheedy, J. E., & McCarthy, M. (1994). Reading Performance and Visual Comfort with Scale to Gray Compared with Black-and-White Scanned Print. *Displays, 15*(1), 27-30.

Sheedy, J. E., Tai, Y.-C., Subbaram, M., Gowrisankaran, S., & Hayes, J. (2008). ClearType subpixel text rendering: Preference, legibility and reading performance. *Displays, 29*(2), 138-151. doi: 10.1016/j.displa.2007.09.016

Shieh, K.-K., & Lin, C.-C. (2000). Effects of screen type, ambient illumination, and color combination on VDT visual performance and subjective preference. *International Journal of Industrial Ergonomics, 26*(5), 527-536. doi: 10.1016/S0169-8141(00)00025-1

Slattery, T. J., & Rayner, K. (2010). The influence of text legibility on eye movements during reading. *Applied Cognitive Psychology, 24*(8), 1129-1148. doi: 10.1002/acp.1623

Smith, M. J., Cohen, B. G. F., Stammerjohn, L. W., & Happ, A. (1981). An investigation of health complaints and job stress in video display operations. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 23*(4), 387-400. doi: 10.1177/001872088102300402

Sotoyama, M., Jonai, H., Saito, S., & Villanueva, M. B. G. (1996). Analysis of ocular surface area for comfortable VDT workstation layout. *Ergonomics, 39*(6), 877-884. doi: 10.1080/00140139608964508

Starr, S. J. (1983). A study of video display terminal workers. *Journal of Occupational and Environmental Medicine, 25*(2), 95-97.

Steyer, R., Schwenkmezger, P., Notz, P., & Eid, M. (1997). *Der mehrdimensionale Befindlichkeitsfragebogen (MDBF) [The multidimensional mood questionnaire (MDBF)]*. Göttingen: Hogrefe.

Thorud, H. M. S., Helland, M., Aarås, A., Kvikstad, T. M., Lindberg, L. G., & Horgen, G. (2012). Eye-related pain induced by visually demanding computer work. *Optometry and Vision Science, 89*(4), 452-464. doi: 10.1097/OPX.0b013e31824c1801

Tsubota, K., & Nakamori, K. (1993). Dry eyes and video display terminals. *New England Journal of Medicine, 328*(8), 584. doi: 10.1056/NEJM199302253280817

Wilkinson, R. T., & Robinshaw, H. M. (1987). Proof-reading: VDU and paper text compared for speed, accuracy and fatigue. *Behaviour and Information Technology, 6*(2), 125-133. doi: 10.1080/01449298708901822

Wright, P., & Lickorish, A. (1983). Proof-reading texts on screen and paper. *Behaviour and Information Technology, 2*(3), 227-235. doi: 10.1080/01449298308914479

Ziefle, M. (2001). Aging, visual performance and eyestrain in different screen technologies. *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting*, 262-266. doi: 10.1177/154193120104500315

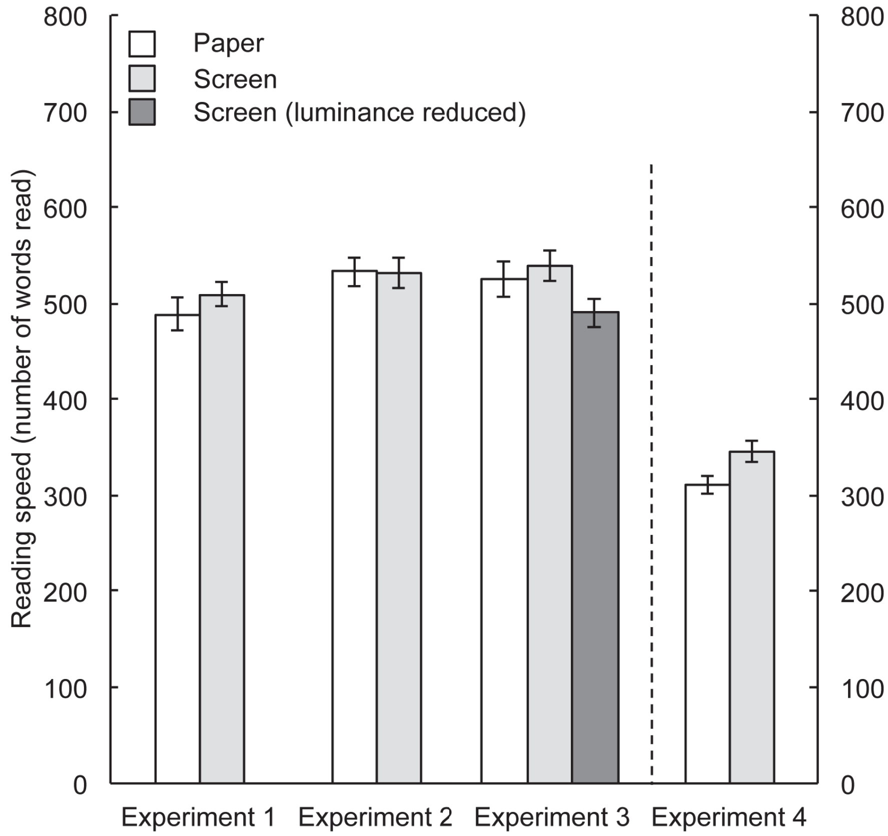
Figure Captions

Figure 1: Mean reading speed (number of words read) as a function of display mode for Experiments 1-3 (3 minutes reading time per text) and Experiment 4 (2 minutes reading time per text). Error bars represent the standard errors of the means.

Figure 2: Mean proofreading performance (number of errors detected – false alarms) as a function of display mode for Experiments 1-3 (3 minutes reading time per text) and Experiment 4 (2 minutes reading time per text). Error bars represent the standard errors of the means.

Figure 3: Experiment 1: Mean for the eyestrain dimension for the first testing. Larger scores indicate a stronger degree of eyestrain; the highest possible symptom score was 49. The error bars represent the standard errors of the means. Experiments 2, 3, and 4: Mean post-pre differences for the eyestrain dimension. Positive scores indicate an increase of eyestrain at the post-measurement; the highest possible difference in symptom score was 42. The error bars represent the standard errors of the mean differences.

Figure 1



800
700
600
500
400
300
200
100
0
Paper
Screen
Screen (luminance reduced)
Reading speed (number of words read)
Experiment 1
Experiment 2
Experiment 3
Experiment 4

Figure 2

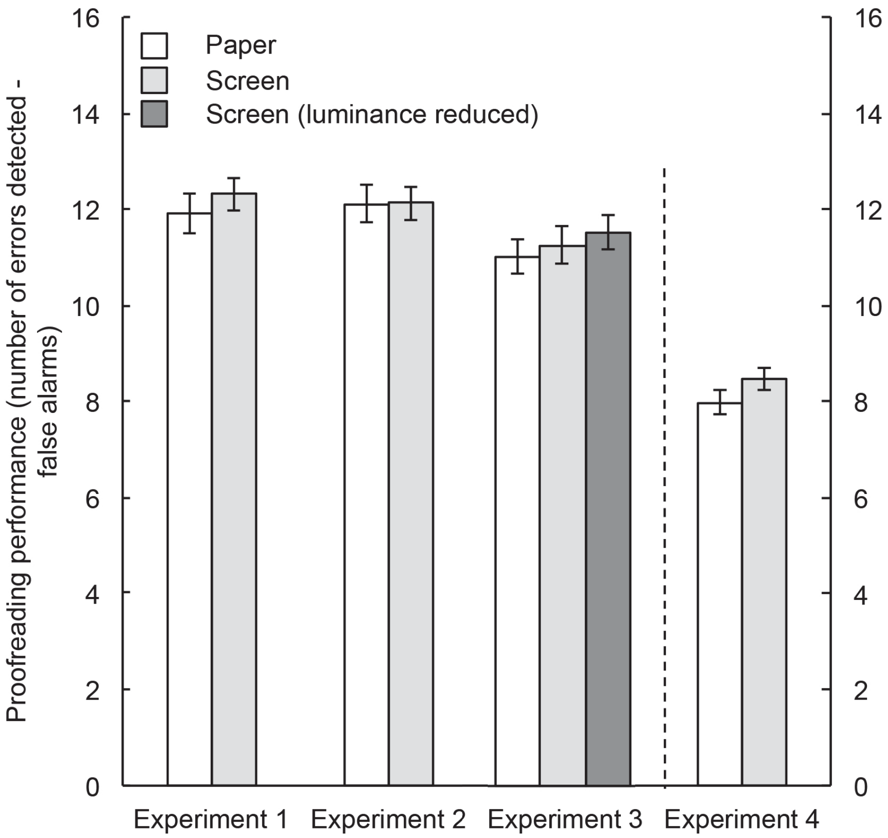
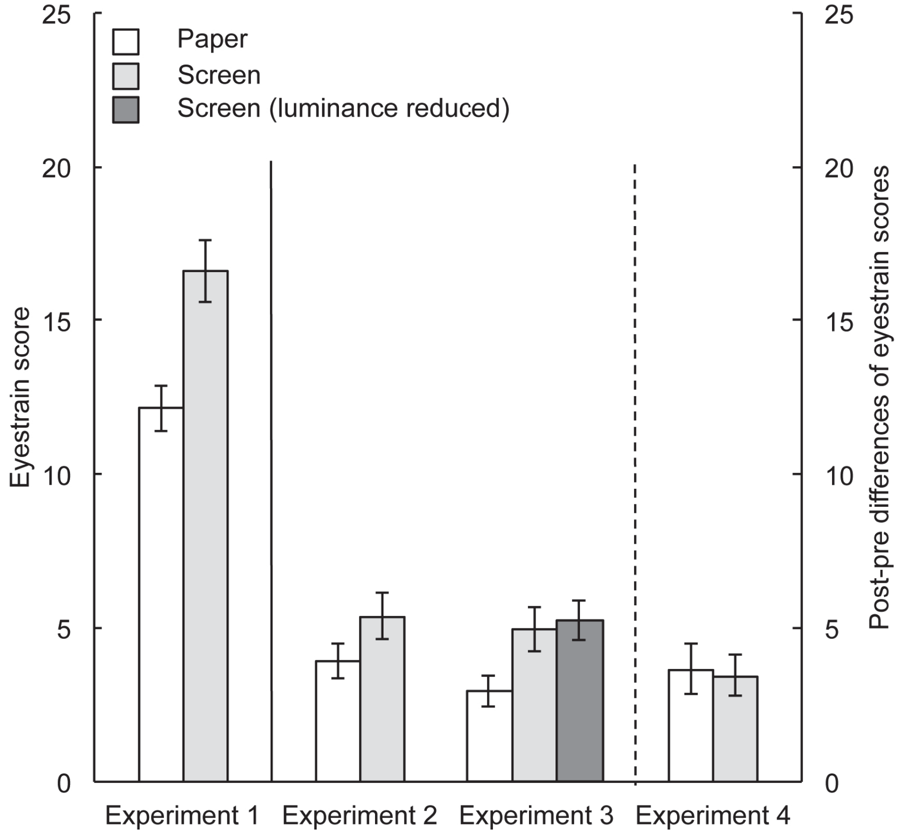

Proofreading performance (number of errors detected - false alarms)
0
2
4
6
8
10
12
14
16
Paper
Screen
Screen (luminance reduced)
Experiment 1
Experiment 2
Experiment 3
Experiment 4

Figure 3


Paper
Screen
Screen (luminance reduced)
Eyestrain score
Post-pre differences of eyestrain scores
0
5
10
15
20
25
Experiment 1
Experiment 2
Experiment 3
Experiment 4

Appendix

In order to get preference judgements based on recent experiences, a sub-sample of n = 30 participants of Experiment 1 (in the following referred to as Experiment 1A) proofread texts from screen and then from paper or vice versa and completed preference questions afterwards. As stated in the introduction to Experiment 1, the results of the second reading condition were not included in the overall statistical analyses because performance on the second of the two tasks is no longer independent, but is contaminated by performance on the first task and is thus likely to exhibit performance-effort trade-offs mentioned in the Introduction section. Therefore, the second reading condition does not provide a usable performance measure. In order to illustrate the possible problems occurring when using within-subjects designs, and the results of the subgroup with two reading tasks in Experiment 1 are briefly presented here.

The means and standard errors of the the means for reading speed, proofreading performance, the three dimensions of the multi-dimensional mood state questionnaire and the three dimensions of the physical discomfort questionnaire are presented in Table A1. For each dependent variable, we conducted a mixed ANOVA (analysis of variance) with display order (screen-paper versus paper-screen) as within-subject variable and display mode (screen versus paper) as between subjects variable. The main effects of display mode and the display order × display mode interactions are presented in Table A2.

The analyses revealed significant main effects of display mode on alertness versus fatigue, eyestrain, and headache (participants reported less fatigue, eyestrain, and headache after reading on paper). However, these main effects of display mode are accompanied by significant display order × display interactions. For example, an equal amount of eyestrain after reading on screen and paper was reported when participants started with reading on screen. In contrast, eyestrain was stronger after reading on screen, when participants read the texts on paper first. Thus, due to influence of display order on the amount of eyestrain symptoms the main effects of display mode on eyestrain cannot be interpreted properly.

Table A1

Means and standard errors of the means of reading speed, proofreading performance, the three dimensions of the multi-dimensional mood state questionnaire (a lower score indicates a more negative rating) and the three dimensions of the physical discomfort questionnaire (larger scores indicate a stronger degree of physical discomfort) as a function of display order and display for Experiment 1A.

	Display order	Display mode	*M*	*SE*
reading speed	screen-paper	paper	541	27
		screen	503	28
	paper-screen	paper	501	29
		screen	515	30
proofreading performance	screen-paper	paper	12.27	0.77
		screen	11.88	0.73
	paper-screen	paper	11.67	0.83
		screen	11.52	0.86
pleasant versus unpleasant mood	screen-paper	paper	15.33	0.56
		screen	15.05	0.71
	paper-screen	paper	16.58	0.59
		screen	15.26	0.74
alertness versus fatigue	screen-paper	paper	12.14	0.71
		screen	12.33	0.72
	paper-screen	paper	14.05	0.75
		screen	10.95	0.75
calmness versus agitation	screen-paper	paper	15.19	0.67
		screen	15.52	0.65
	paper-screen	paper	15.63	0.71
		screen	15.11	0.68
eyestrain	screen-paper	paper	13.62	1.93
		screen	13.43	1.72
	paper-screen	paper	14.84	2.03
		screen	20.84	1.81
headache	screen-paper	paper	5.91	0.89
		screen	5.86	0.96
	paper-screen	paper	5.84	0.94
		screen	8.21	1.01

musculoskeletal strain	screen-paper	paper	8.38	1.12
		screen	7.38	0.99
	paper-screen	paper	6.74	1.18
		screen	8.74	1.04

Note. Display order: screen - paper versus paper - screen. Display mode: screen versus paper.

Table A2

Statistical results of the 2 × 2 mixed ANOVA with display mode as within-subject variable and display order as between-subjects variable for Experiment 1A.

	Test	*F*	*df*	*p*	η^2
reading speed	display mode	3.39	1,37	.07	.08
	display mode × display order	**14.35**	**1,37**	**< .001**	**.28**
proofreading performance	display mode	1.36	1,37	.25	.04
	display mode × display order	0.29	1,37	.59	.01
pleasant versus unpleasant mood	display mode	2.65	1,38	.11	.07
	display mode × display order	1.10	1,38	.30	.03
alertness versus fatigue	display mode	**9.61**	**1,38**	**.004**	**.20**
	display mode × display order	**12.29**	**1,38**	**.001**	**.24**
calmness versus agitation	display mode	0.04	1,38	.85	.001
	display mode × display order	0.72	1,38	.40	.02
eyestrain	display mode	**7.99**	**1,38**	**.007**	**.17**
	display mode × display order	**9.07**	**1,38**	**.005**	**.19**
headache	display mode	**9.568**	**1,38**	**.004**	**.20**
	display mode × display order	**10.37**	**1,38**	**.003**	**.21**
musculoskeletal strain	display mode	0.89	1,38	.35	.02
	display mode × display order	**8.05**	**1,38**	**.007**	**.18**

Note. Display order: screen - paper versus paper - screen. Display mode: screen versus paper. Significant effects are printed in bold.

Effects of high pixel density on reading comprehension, proofreading, and subjective well-being

Maja Köpper, Susanne Mayr & Axel Buchner

Running head: Effects of high pixel density

Abstract

Displays with low pixel densities that were common in the 1980s and 1990s were shown to impair visual performance. Display technology, especially pixel density, has tremendously improved in recent years and new technologies allow densities of 264 ppi and beyond. Two experiments were conducted to test whether high-resolution displays have beneficial effects on visual performance and subjective well-being. In Experiment 1, participants performed a reading comprehension task on a display with high pixel density (264 ppi) or on a display with moderate pixel density (132 ppi). In Experiment 2, participants' speed and performance in a proofreading task were compared using the same two displays. There were no differences in reading comprehension and reading time (Experiment 1), proofreading speed and performance (Experiment 2) between a 132 ppi and a 264 ppi display. However, subjective ratings of physical discomfort revealed significantly more complaints about headache and musculoskeletal strain in the 132 ppi condition than in the 264 ppi condition (Experiment 2). Although reading comprehension and visual performance is unaffected by pixel densities above 132 ppi, reading from high resolution screens seems to be less exhausting. Displays with high pixel densities (264 ppi and above) should be preferred over displays with standard (132 ppi or lower) pixel densities.

Keywords

display resolution, iPad, reading speed, well-being, legibility

1. Introduction

Developments in screen technology and display design that have been introduced over the past years have resulted in improvements of the legibility of text presented on computer screens. This can be concluded from two facts. First, whereas studies from the 1980s and 1990s revealed clear performance detriments when reading from cathode ray tube (CRT) screens as compared with reading from paper (for comprehensive reviews of these earlier studies see e.g., Dillon, 1992; Mills & Weldon, 1987) more recent studies using modern liquid crystal displays (TFT-LCDs) revealed comparable performance for screen and paper (Gujar, Harrison, & Fishkin, 1998; Holzinger et al., 2011; Köpper, Mayr, & Buchner, 2015; Kretzschmar et al., 2013; Margolin, Driscoll, Toland, & Kegler, 2013; Meyer & Poon, 1997; Moore & Zabrucky, 1995). Second, direct comparisons between TFT-LCDs and CRTs have repeatedly revealed a superiority of TFT-LCDs over CRT screens with regard to measures of visual performance such as visual search speed, search accuracy, letter identification, and eye movement parameters (Menozzi, Lang, Näpflin, Zeller, & Krueger, 2001; Menozzi, Näpflin, & Krueger, 1999; Näsänen, Karlsson, & Ojanpää, 2001; Shieh & Lin, 2000; Ziefle, 2001).

Among other variables (such as a flicker-free presentation, high background luminance and luminance contrast, and improved text presentation capabilities), increased pixel density is one factor that accounts for the improved legibility of text on modern TFT-LCDs. Pixel density is defined as the number of pixels per length unit (inch) of a display (pixel per inch; ppi) and is often mistakenly equated with display resolution that refers to the total number of pixels in each dimension of the display. Note that two displays with different display resolutions may have similar pixel densities if they have different display sizes and vice versa. In the following, we use display resolution as a synonym for pixel density only if the display sizes are equivalent.

The increase in display sharpness that comes along with increases in pixel density leads to improved legibility of text presented on CRT screens. For example, in a post-hoc analysis of several previous experiments Gould, Alfaro, Finn, Haupt, and Minuto (1987) showed an influence of pixel density on proofreading. Specifically, proofreading speed increased as a function of the number of pixels per inch (within a range from 62 ppi to 92 ppi). Ziefle (1998, Experiment 2) examined the effects of different pixel densities (62, 69, and 89 ppi) on visual search, eye-movement parameters, and visual fatigue. She showed slower visual search times and longer fixation durations in the 62 ppi condition than in the 89 ppi condition. There was no significant effect of pixel density on ratings of visual fatigue. In another experiment of the same study (Ziefle, 1998, Experiment 1), there was no difference in proofreading performance between a CRT screen with a pixel density of 60 ppi and one with 120 ppi. However, participants' ratings indicated a preference for the 120 ppi condition (although a paper condition with an even higher pixel density of the printed letters [255 ppi] was preferred by a large majority of the participants). Bridgeman, Lennon, and Jackenthal (2003) compared reading comprehension on a high-resolution display (17 inch, 1024 × 768 pixels) and a low resolution display (17 inch, 640 × 480 pixels). The comprehension scores were numerically higher in the higher resolution condition, but the difference missed statistical significance.

Pixel density has been increased tremendously over the past 30 years: Whereas pixel densities within a range of 60-90 ppi were common 30 years ago (e.g., Gould et al., 1987), pixel densities of up to 120 ppi (e.g., Ziefle, 1998) were common in the late 1990ies and at the beginning of the 21st century. In 2010 and 2011, Apple introduced the first and second generations of their iPad with a 9.7 inch and 132 ppi (1024 × 768 pixels) display, followed by the iPad 3 in 2012 with a 9.7 inch display with a pixel density of 264 ppi (2048 × 1536 pixels, socalled retina display). By now, pixel densities at this level and higher have become the standard for most mobile phones and tablet computers as well as for some laptop displays.

The pixel density of 264 ppi of the iPad 3 results in a pixel size of 0.096 mm (~ 96 µm). The "normal" human eye can discriminate two pixels separated by a gap of 1 arcminute (1 arcminute is the size of the gap of a standard Landolt C at a viewing distance of 20 feet or 6 m that equals a visual angle of about 0.017°), although the standard visual acuity of 20/20 (1.0) is often exceeded by young and healthy people or when defective vision is well-corrected. The visual angle of one pixel presented at the iPad 3 equals 1 arcminute when the viewing distance is set to approximately 33 cm. At this distance and beyond, single pixels can no longer be distinguished from another by a "normal" human eye. Thus, the goal of display research "to match the information output of the display to the information capacity of the human visual system" (Alt & Noda, 1998, p. 315) seems to have been achieved.

However, whereas the influence of pixel density on legibility has been shown for CRT screens, the pixel densities in those studies were considerably lower than the number of pixels per inch that come along with state-of-the-art LCD displays. Studies investigating the beneficial effects of high resolution displays with pixel densities at or even beyond the limit of the human eye's resolution capability are rare. One exception is the study reported by Gujar et al. (1998) who compared an LCD with high pixel density (282 ppi), a CRT display (85 ppi), and print on paper (300 ppi). The authors used a short modified proofreading task that focused on comprehension rather than on scanning the text for misspellings and found no significant differences in reading speed, error rates, and viewing distance. Gujar et al. pointed out that existing differences in legibility may have been masked by the short duration and the low complexity of their task. However, participants rated the ease of reading as highest in the paper condition as compared with the screen conditions. The 85 ppi CRT screen was rated significantly worse than the other conditions. Wright, Bailey, Tuan, and Wacker (1999) compared a 102 ppi CRT display with an 83 ppi and a 157 ppi TFT-LCD. They found no differences in visual search performance, reading speed, and comprehension among the displays, but legibility in terms of visual acuity (number of words read in a word chart) increased with increasing pixel density. Further, participants preferred the 157 ppi TFT-LCD over the other displays. The 83 ppi TFT-LCD was the least preferred display and the worst in overall visual comfort. The authors also tested an additional condition in which the CRT display was positioned at a viewing distance that resulted in the same retinal size of the pixels as the 157 ppi TFT-LCD condition. Interestingly, there were no differences in preference and subjective ratings of overall visual comfort between these two conditions, indicating that retinal pixel size and, consequently, pixel density are determining factors in subjective evaluations of display quality. More recently, Huang, Rau, and Liu (2009) found slower reading speed for a pixel density of 125 ppi as compared with densities of 167 ppi, 200 ppi, and 250 ppi. Visual search performance was not affected by pixel density. The authors concluded that higher pixel densities can provide higher legibility. Note that participants read texts or conducted a visual search task on different mobile devices with various screen and font sizes. Thus, the manipulation of pixel density was confounded with these variables. For instance, the largest font sizes were used in the 125 ppi condition, most probably resulting in longer reading times due to longer scrolling times.

Although empirically the effects of pixel density on measures of performance, well-being, and preference seem somewhat inconsistent, a cautious but acceptable conclusion seems to be that an increase in pixel density form about 60 ppi to about 150 ppi has positive effects (as shown by Gould et al., 1987; Huang et al., 2009; Wright et al., 1999; Ziefle, 1998). The purpose of the experiments presented here was to test whether and increase in pixel density form 132 ppi to 264 ppi would continue to have positive effects on measures of visual performance and subjective well-being. In Experiment 1, we investigated possible effects of pixel density in a reading-for-comprehension task, whereas in Experiment 2 a proofreading task was used that would require focussing more closely on the perceptual details of the text.

2. Experiment 1

We measured participants' reading comprehension and reading time using two displays that differed in pixel density but were basically identical with respect to all crucial display variables such as display size, luminance, luminance contrast, and anti-aliasing algorithms. Reading comprehension was chosen as the dependent variable because this task is often required in the daily routine as well as at work (e.g., reading news online, e-mails, safety instructions, e-learning etc.) and therefore provides a measure of high ecological validity.

It is conceivable that participants perceive impairments of their performance, relative to the very familiar reading-from-paper situation, in the more difficult reading-from-screen condition (which might be the condition with the lower pixel density in the present experiment) and, in turn, try to maintain a certain level of performance over the course of the experiment. First, participants may try to adjust their reading speed in order to maintain a constant level of comprehension (e.g., Dillon, 1992; Mills & Weldon, 1987). This assumption is supported by findings of Cushman (1986) who showed a negative correlation between reading speed and comprehension. Hence, an effect of pixel density on reading comprehension may be reflected partially, perhaps even completely, in the time taken to read the texts. Therefore, we measured reading time in addition to text comprehension. Second, participants may attempt to compensate the adverse effects of a lower pixel density on comprehension and reading speed by increased effort (Buchner & Baumgartner, 2007; Ziefle, 1998). Thus, in order to assess how much participants strained themselves, the performance measurements were complemented by subjective ratings of psychological and physical well-being. Finally, participants may try to compensate for legibility deficits by increasing or decreasing their viewing distance. For instance, the lower pixel density might cause a blurred display image that participants might compensate by reducing their viewing distance. Alternatively, an increased viewing distance causes a smaller retinal image of the pixels and might increase the perceived image quality (Wright et al., 1999). Such adjustments are possible everyday tasks in which people are usually free to choose their preferred seating position when reading text from a screen. Restricting participants' posture (e.g, by the use of chin rests) would most probably result in a severe reduction of ecological validity. Therefore, we did not restrict the seating position of the participants. Instead we measured the chosen viewing distance as an additional indicator of perceived visual quality at the end of the experiment.

2.1. Method

2.1.1. Participants

Participants were 156 adults. One data set had to be excluded from the analyses because, by accident, the participant was given a non-corresponding combination of texts and comprehension questions. Thus, the final sample size was 155. The 132 ppi condition comprised 80 participants (14 male) who ranged in age from 18 to 40 years (M = 22). The 264 ppi condition comprised 75 participants (20 males) who ranged in age from 18 to 32 years (M = 22). The groups did not differ with regard to their familiarity with the texts (Question 1 of the final questionnaire; for details see below), $\chi^2(1) = 0.01$, $p = 1.00$, their familiarity with the task (Question 2), $\chi^2(1) = 0.21$, $p = .77$, and their reading time per day (Question 8), $\chi^2(5) = 9.23$, $p = .08$. Similarly, there were no differences between the groups regarding the use of an iPad or another tablet computer (Question 5), $\chi^2(1) = 0.03$, $p = 1.00$, the use of an iPod touch, iPhone, or another smartphone (Question 6), $\chi^2(1) = 0.22$, $p = .75$, or their experience with those devices (Question 7), $\chi^2(2) = 1.14$, $p = .60$. All participants except one were students. They were either paid or received partial course credit for their participation. All participants were native German speakers, reported normal or corrected-to-normal visual acuity, and had not been diagnosed with dyslexia. One participant decided not to fill out any of the questionnaires, one participant did not answer the multi-dimensional mood state questionnaire, one participant did not answer the physical discomfort questionnaire, and another participant did not answer the musculoskeletal pain questions of the physical discomfort questionnaire. For three participants the viewing distance measurements were missing. The missing values imply varying degrees of freedom for the different analyses reported.

2.1.2. Material and Task

The experiment took place in two identical sound-insulated testing chambers without any exposure to daylight. Each chamber was furnished with a 100 × 100 × 71.5 cm (width × length × height) table and a height-adjustable office chair. The displays were positioned on a desk stand with a 15° inclination angle relative to the horizontal axis in order to enable a comfortable reading position (Bridger, 1988; Eastman & Kamon, 1976). The mostly indirect ambient room lighting was produced by a fluorescent lamp (4000 K correlated colour temperature). Ambient illumination at the participants' approximate eye position (measured with a Gossen Mavolux 5032 B illuminance meter; Gossen Foto- und Lichtmesstechnik GmbH, Nürnberg, Germany; with Class B accuracy according to DIN 5032-7) was 620 lx in the 132 ppi condition and 577 lx in the 264 ppi condition. Luminance values and luminance contrasts (see Table 1) were determined using a Minolta Colormeter CS-100 (Konica Minolta Co. Ltd, Tokyo, Japan).

Table 1

Luminance values (cd/m^2) and luminance contrasts of the 132 ppi and the 264 ppi display. As luminance measurements in the two chambers differed only marginally—on average, they varied by 7.04 % (SD = 5.68)—average values are reported.

	Display luminance (cd/m^2)	Luminance contrast[a]
132 ppi		
black	6	0.96
white	342	
264 ppi		
black	6	0.97
white	383	

[a] Luminance contrast = $(L_{max} - L_{min})/(L_{max} + L_{min})$

The reading comprehension task comprised 12 individual short stories (travelogues and economic contents) by various authors. The stories were presented in random order. After reading a text, participants answered five content-related single-choice and multiple-choice questions.

Texts were typeset in LaTeX (KOMA script class scrartcl) and presented as PDF pages in Apple iBooks 2 running under iOS 5.1. In the 132 ppi condition texts were presented on the 9.7" (1024 × 768 pixel) glossy widescreen Multi-Touch display with LED-backlight and IPS technology of an Apple iPad 2. In the 264 ppi condition a 9.7" (2048 × 1536 pixel) glossy widescreen Multi-Touch retina display with LED-backlight and IPS technology of an Apple iPad 3 was used. All texts were anti-aliased.

The mean length of the texts was 580 words (SD = 22). The texts were presented in 12 pt Helvetica font (x-height = 2.1 mm, cap-height = 2.8 mm), single-spaced. Texts were presented in horizontal format and consisted of two columns with a height of 14.1 cm. The columns were typeset in justified format with hyphenation. Each column comprised a total of 34 lines that were 9 cm long. The text layout was exactly the same for both conditions.

The reading comprehension test was presented on an Apple MacBook Pro computer (15.4", 1680 × 1050 pixel, 128 ppi, TFT-LCD widescreen display with LED-backlight) that also controlled the experiment. The computer was positioned right behind the desk stand on the table about 80 cm away from the participants. Five content-related questions were created for each text. Each question comprised five response options. We used both single-choice items (only one correct option, e.g. “In which city did the reporters land?”) and multiple-choice items (one or more correct options, e.g. “What is in the one-room cottage?”). Participants knew at least one answer was correct, but they did not know the exact number of correct answers for the multiple-choice items. To reach the next item, participants had to choose at least one option.

For each question, a score was computed. For single-choice questions the score was 1 if the correct option was chosen, and 0 otherwise. For multiple-choice items every correctly chosen or correctly rejected option + 0.2 was added to the score, and 0.2 was subtracted for each mistakenly chosen or rejected option. Accordingly, the highest possible score was 1. The lowest score was 0 (negative scores were treated as 0). The average scores for the final sample ($N = 155$) ranged between 0.35 and 0.89 ($M = 0.68$, $SD = 0.13$). Cronbach’s α was .86.

A pilot test had revealed that at least 2 minutes were necessary to read the texts for comprehension. Therefore, reading times faster than 90 seconds were considered unrealistically short and most probably caused by skimming rather than reading the texts. Thus, only texts with a reading time longer than 90 seconds were included in the analysis.

The two parallel short forms of the multi-dimensional mood state questionnaire (Steyer, Schwenkmezger, Notz, & Eid, 1997) were used to measure the participants’ mood state on three bipolar dimensions of psychological well-being: pleasant versus unpleasant mood, alertness versus fatigue, and calmness versus agitation. The 24 items (12 in each short form) comprised 24 simple adjectives (e.g., tired, relaxed) as descriptions of the participants’ current mood state, which the participants rated on a 5-point scale from “not at all” (1) to “very much” (5).

In the physical discomfort questionnaire, which was based on the questionnaire developed by Heuer, Hollendiek, Kröger, and Römer (1989), participants rated several physical symptoms of strain. The questionnaire comprised 14 statements about symptoms of eyestrain (7 items), headache (3 items), and musculoskeletal strain (4 items). Participants rated whether they experienced the symptoms on a 7-point scale from “not at all” (1) to “very much” (7).

A final questionnaire was used to assess the possible influence of several additional variables. First, the short stories that we used in Experiment 1 and 2 can be found on the internet and the proofreading task in Experiment 2 is a frequently used experimental task. Participants might have felt familiar with the task or some of the texts and therefore were asked whether they remembered the texts or the task from another context (Questions 1 and 2, respectively). Further, participants were requested to assess reading comfort on a 5-point scale, such as the perceived contrast of the display (too low [1] to too high [5]; Question 3), the perceived brightness of the display (too low [1] to too high [5]; Question 4), the quality of their reading position (very uncomfortable [1] to very comfortable [5], Question 5), the impairment by glare (not at all [1] to very much [5]; Question 6), and the impairment by reflections (not at all [1] to very much [5]; Question 7). Finally, participants were asked how much time they spent on reading per day (options were: up to 30 minutes, up to 1 hour, up to 2 hours, up to 4 hours, up to 6 hours, more than 6 hours; Question 8). There was also free space for any other comments on the experiment. All questionnaires were printed on paper.

Note that an additional paper condition was tested in the original Experiment. Therefore, questions concerning the comparison of reading on screen and paper (e.g., the preference for paper or screen in daily routine) were asked. These questions are irrelevant for the comparison of the different display resolutions and therefore are not reported.

2.1.3. Procedure

Participants were tested individually. Before starting the experiment they were asked to set the height-adjustable seat to their preferred position. Subsequently, participants were randomly assigned to the 132 or to the 264 ppi condition.

The participants first answered one of the two parallel short forms of the multi-dimensional mood state questionnaire and the physical discomfort questionnaire. Subsequently, the main instructions were presented on the laptop screen. The general instructions explaining the reading task and the comprehension test were followed by a short introduction to the handling of the iPad. There was also a short practice trial in order to familiarise the participants with the procedure of the experiment. At the beginning of each trial, participants started the measurement of the reading time by clicking on the "start reading" button on the laptop screen. Subsequently, they were auditorily prompted to open the display cover (Apple iPad smart cover), to turn over to the first page, and to start reading. Participants were instructed to read and reread the text until they felt familiar with the content and to stop the reading time by clicking on the "stop reading" button on the laptop screen as soon as they had finished reading. They were subsequently asked to close the display cover in order to hide the text during the comprehension test on the laptop. Participants responded to the comprehension questions at their own pace. Subsequently, they moved on to read the next text on the screen. This procedure was identical for all 12 texts and their associated comprehension questions. At the end of the experiment when the participants were familiar with the experimental task and had found their favourite reading position, they were auditorily cued to stop reading after 10 seconds of reading an additional (thirteenth) text, and to keep their current reading position until the experimenter had measured the distance between the participant's eye and the display with a meter stick. Finally, the remaining parallel short form of the multi-dimensional mood state questionnaire, the physical discomfort questionnaire, and the final questionnaire were administered. The entire experimental procedure took about 60 to 90 minutes depending on the individual reading speed of the participants.

2.1.4. Design

The experiment comprised a one-factorial design with pixel density (132 vs. 264 ppi) as between-subjects variable. The dependent variables were participants' mean comprehension score, reading time (in seconds), the three mood state scores, the physical discomfort scores of eyestrain, headache, and musculoskeletal strain, and the viewing distance between participants' eyes and the display of the tablet computer.

Assuming a medium to large effect size of $d = 0.60$ of the pixel density manipulation (as defined by Cohen, 1988), and desired levels of $\alpha = \beta = .05$, an a priori power analysis revealed that data would have to be collected from $N = 122$ participants (Faul, Erdfelder, Lang, & Buchner, 2007) We were able to collect data from $N = 155$ participants so that the power was even larger than what we had planned for ($1 - \beta = .98$). The level of alpha was maintained at .05 for all statistical decisions. Corresponding to the directional hypotheses, all reported *t* tests are one-sided, with the exception of the reading comfort evaluations that are based on two-tailed *t* tests.

2.2. Results

2.2.1. Mean comprehension score and reading time

The mean comprehension score was computed by summing the scores for each text and averaging over all texts for each participant. When computing the mean reading times, data from 0.75 % of the texts were dropped due to implausibly short measurements (see Materials and Task section). Independent groups *t* tests showed no effect of pixel density on the mean comprehension score, $t(153) = 1.16$, $p = .12$, $d = 0.19$, and reading time, $t(153) = 0.91$, $p = .18$, $d = 0.15$ (Figure 1).

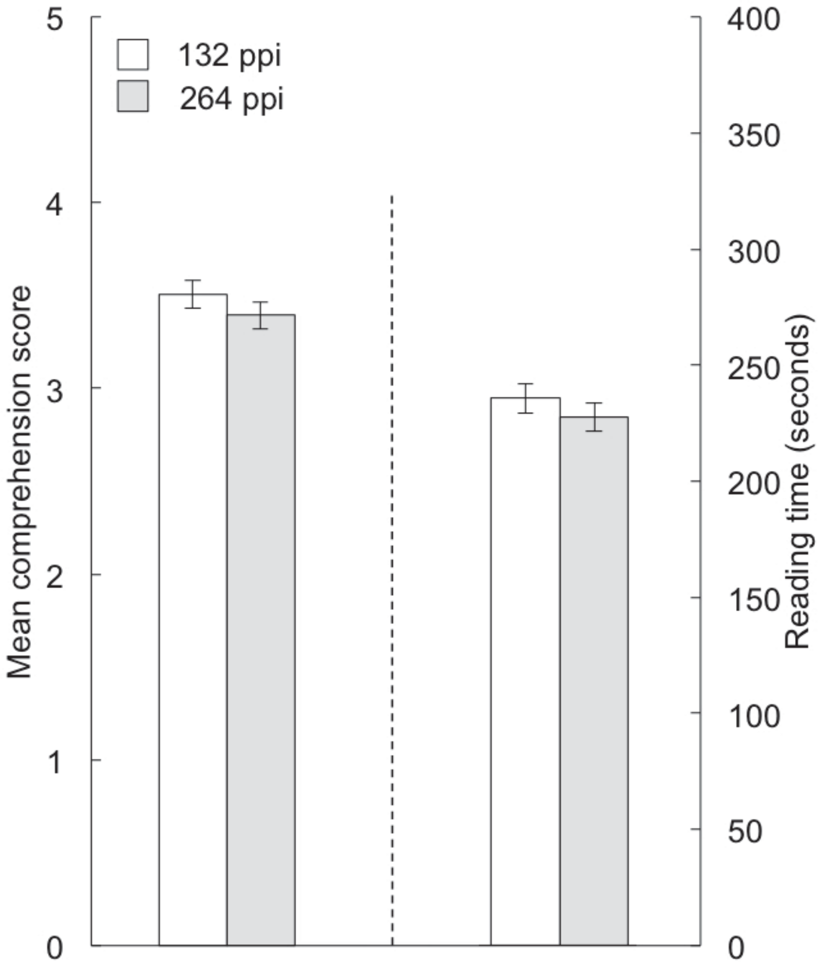


Figure 1

Mean comprehension score on the left panel and mean reading time (seconds) on the right as a function of pixel density for Experiment 1. Error bars represent the standard errors of the means.

2.2.2. Mood state and physical discomfort ratings

The mean differences between post- and pre-values for all dimensions of the multi-dimensional mood state questionnaire and the physical discomfort questionnaire are presented in Tables 2 and 3, respectively. There was a descriptive trend towards a stronger decrease of mood (pleasant vs. unpleasant) in the 132 ppi than in the 264 ppi condition, but this trend was not statistically significant, $t(151) = 1.34$, $p = .09$, $d = 0.22$. Pixel density did not affect changes in participants' ratings in the remaining "alertness versus fatigue" and "calmness versus agitation" dimensions, $ts(151) < 0.58$, $ps > .28$, $ds < 0.09$.

Table 2

Mean post-pre differences of the three dimensions of the multi-dimensional mood state questionnaire and standard errors of the mean differences for Experiment 1 and Experiment 2. Negative scores indicate worse mood ratings at the post-measurement.

	pixel density	*N*	*M*	*SE*
Experiment 1				
pleasant versus unpleasant mood	132 ppi	79	-1.18	0.28
	264 ppi	74	-0.65	0.28
alertness versus fatigue	132 ppi	79	-2.63	0.42
	264 ppi	74	-2.65	0.36
calmness versus agitation	132 ppi	79	-0.65	0.36
	264 ppi	74	-0.91	0.26
Experiment 2				
pleasant versus unpleasant mood	132 ppi	76	-0.20	0.29
	264 ppi	78	-0.06	0.24
alertness versus fatigue	132 ppi	76	-1.33	0.28
	264 ppi	78	-2.01	0.30
calmness versus agitation	132 ppi	76	-0.53	0.28
	264 ppi	78	-0.03	0.30

Similarly, there was a descriptive trend towards a stronger increase of eyestrain in the 132 ppi than in the 264 ppi condition, but this trend also just missed the preset level of significance, $t(151) = 1.50$, $p = .07$, $d = 0.24$. Pixel density did not affect ratings of headache, $t(151) = 0.58$, $p = .29$, $d = 0.09$, or musculoskeletal strain, $t(150) = 1.18$, $p = .12$, $d = 0.19$.

Table 3

Mean post-pre differences of the three dimensions of the physical discomfort questionnaire and standard errors of the mean differences for Experiment 1 and Experiment 2. Positive scores indicate an increase of physical discomfort at the post-measurement.

	pixel density	*N*	*M*	*SE*
Experiment 1				
eyestrain	132 ppi	79	4.22	0.65
	264 ppi	74	2.76	0.73
headache	132 ppi	79	1.76	0.32
	264 ppi	74	2.01	0.30
musculoskeletal pain	132 ppi	78	1.56	0.31
	264 ppi	74	1.07	0.29
Experiment 2				
eyestrain	132 ppi	75	2.87	0.57
	264 ppi	77	2.21	0.51
headache	132 ppi	75	1.91	0.35
	264 ppi	78	1.21	0.23
musculoskeletal pain	132 ppi	75	2.31	0.47
	264 ppi	78	1.21	0.34

2.2.3. Viewing distance

Participants chose a similar viewing distance in the 132 ppi condition (*M* = 39.28 cm, *SE* = 0.69 cm) and in the 264 ppi condition (*M* =38.41 cm, *SE* = 0.59 cm), $t(150) = 0.96$, $p = .17$, $d = 0.16$.

2.2.4. Evaluations of reading comfort and preference judgement

There were no significant differences between the 132 ppi and the 264 ppi condition in evaluations of perceived contrasts, $t(151) = 0.27$, $p = .40$, $d = 0.04$, brightness, $t(152) = 0.28$, $p = .39$, $d = 0.05$, comfort of reading position, $t(152) = 0.42$, $p = .34$, $d = 0.07$, impairment by glare, $t(152) = 1.66$, $p = .05$, $d = 0.27$, and impairment by reflections $t(152) = 1.52$, $p = .07$, $d = 0.25$ (Question 3 to 7; Table 4), with the latter two comparisons just missing the preset level of statistical significance.

Table 4

Mean values and standard errors of the means of the evaluations of contrast, brightness, position, glare, and reflections for Experiment 1 and Experiment 2.

	pixel density	*N*	*M*	*SE*
Experiment 1				
contrast	132 ppi	80	3.16	0.07
	264 ppi	73	3.14	0.06
brightness	132 ppi	80	3.46	0.08
	264 ppi	74	3.43	0.07
position	132 ppi	80	3.49	0.09

		264 ppi	74	3.54	0.08
	glare	132 ppi	80	2.29	0.13
		264 ppi	74	2.00	0.11
	reflections	132 ppi	75	1.61	0.12
		264 ppi	77	1.39	0.08
Experiment 2					
	contrast	132 ppi	76	3.11	0.08
		264 ppi	76	3.16	0.06
	brightness	132 ppi	77	3.44	0.07
		264 ppi	78	3.35	0.07
	position	132 ppi	76	2.99	0.10
		264 ppi	79	3.10	0.10
	glare	132 ppi	77	2.14	0.12
		264 ppi	79	2.08	0.12
	reflections	132 ppi	77	1.56	0.12
		264 ppi	77	1.43	0.09

3. Discussion

The purpose of Experiment 1 was to test whether an increase in pixel density from 132 ppi to 264 ppi would lead to a measurable improvement in reading comprehension or reading time. This was clearly not the case. There were also no differences in the chosen viewing distance and in the measures of subjective well-being. Descriptively, participants' mood and eyestrain ratings were somewhat worse in the 132 ppi than in the 264 ppi conditions, but these differences were small and not statistically significant. Note, that the present sample size allowed to reveal a medium-to-large effect of pixel density with a statistical power of .95. We thus conclude that any differences between the screens with different pixel densities are probably small and of little practical relevance.

Ziefle (1998) already presumed that CRT pixel densities above a certain level might not provide further legibility improvements and therefore might not result in improved visual performance. Similarly, Gujar et al. (1998) showed no proofreading speed or accuracy differences between conditions with 85 ppi and 282 ppi pixel densities, and Huang et al. (2009) found no further reading speed improvements above 167 ppi. The results of Experiments 1 seem generally consistent with those findings.

Alternatively, the comprehension measure might not be sensitive enough to reflect existing differences in legibility due to pixel density. As compared with proofreading, visual search, or visual detection tasks, reading comprehension may be less dependent on perceptual details. For instance, a high comprehension score is based not only on text processing, but also on recognition memory of the text content, and the variance added by recognition memory processes might have masked any differences in visual performance due to differences in the display quality. Gujar et al. (1998) argued that post-reading comprehension tests may result in skimming rather than reading a text, resulting in highly variable reading times and error rates that complicate comparisons between conditions and, therefore, recommended using proofreading tasks to assess the effects of display mode and quality on performance. Proofreading requires both the recognition of perceptual details as word and letter shapes to detect misspellings and comprehension to identify grammar errors. Additionally, the proofreading task has frequently been shown to be sensitive to reading performance differences between screen and paper (e.g., Creed, Dennis, & Newstead, 1987; Gould et al., 1987; Gould & Grischkowsky, 1984; Wilkinson & Robinshaw, 1987) and display polarity (Buchner & Baumgartner, 2007; Buchner, Mayr, & Brandt, 2009; Mayr & Buchner, 2010; Piepenbrock, Mayr, & Buchner, 2014a, 2014b; Piepenbrock, Mayr, Mund, & Buchner, 2013). Furthermore, proofreading is a frequent real-life task, providing a high amount of ecological validity. Accordingly, the aim of Experiment 2 was to test whether proofreading speed and performance would show an advantage of the high pixel density condition.

4. Experiment 2

4.1. Method

4.1.1. Participants

Participants were 161 adults. Two participants had difficulties to do the task properly, causing two times slower reading times and eight times more false alarms than the group average. Two further participants deliberately ignored the instructions (they skipped texts and named words that did not exist in the texts). For one participant, the sound files could not be analysed due to unclear pronunciation. Data of these five participants were excluded from the analyses. The final sample size comprised 156 participants. The 132 ppi condition comprised 77 participants (22 males) who ranged in age from 17 to 37 years ($M = 23$). The 264 ppi condition comprised 76 participants (19 males) who ranged in age from 18 to 40 years ($M = 23$). The groups did not differ with regard to the participants' familiarity with the texts (Question 1 of the final questionnaire; for details see below), $\chi^2(1) = 0.01$, $p = 1.00$, their familiarity with the task (Question 2), $\chi^2(1) = 2.34$, $p = .19$, and their reading time per day (Question 9), $\chi^2(5) = 3.15$, $p = .70$. Similarly, there were no differences between the groups regarding the use of an iPad or another tablet computer (Question 11), $\chi^2(1) = 0.58$, $p = .59$, and the use of an iPod touch, iPhone, or another smartphone (Question 12), $\chi^2(1) = 0.04$, $p = .87$. The groups differed with regard to their experience with those devices (Question 13), $\chi^2(2) = 7.22$, $p = .028$.

However, there were no significant correlations between user experience and the dependent variables (all *p*s > .16) and hence, an influence of the group differences on the present results seems unlikely.

All participants except four were students. They were either paid or received partial course credit for their participation. All participants were native German speakers, reported normal or corrected-to-normal visual acuity, and had not been diagnosed with dyslexia. None of them had participated in Experiment 1. Two participants did not fill in any of the questionnaires, one participant did not answer the physical discomfort questionnaire, and another participant did not respond to the eyestrain questions. For two participants the viewing distance measurements were missing.

4.1.2. Material and Task

Materials and task were the same as those of Experiment 1 with the following exceptions. The proofreading task comprised 20 individual short stories by various authors. Each participant proofread all short stories in the condition he or she was assigned to (132 ppi or 264 ppi). The texts were presented in random order. As in Experiment 1, the texts were presented in horizontal format in two columns, but in contrast to Experiment 1, lines were numbered on both sides of each column. On average, the texts were about 539 (*SD* = 24) words long. Line length was 8.4 cm and the text height was 14.1 cm. The layout was exactly the same for both conditions.

Each story contained 12 misspellings (3 duplicate letters, 3 missing letters, 3 pairwise letter inversions, 3 incorrect letters) and 8 grammar errors (4 incorrect flexions, 4 incorrect conjugations). The grammar errors forced participants to read for comprehension rather than to skim the text for unusual word shapes (Buchner & Baumgartner, 2007; Creed et al., 1987). Errors were randomly distributed across the texts.

Participants were asked to read each text as accurately and fast as possible. Misspellings and grammar errors had to be pronounced along with line number in which the error was identified. Every text was presented for 2 minutes. An auditory signal informed participants when half of the reading time was over. At the end of the two-minute interval, an auditory signal informed participants that the reading time was over and required them to pronounce the word that they had just read and its corresponding line number. Another auditory signal informed participants to turn the page so that they could read the subsequent text.

4.1.3. Procedure

The procedure was identical to that of Experiment 1, with the following exceptions. The instructions for the proofreading task informed participants of the different types of errors and the verbal response mode. They received a short practice trial in which they proofread a training text for 50 seconds in their assigned display resolution condition. The experimenter checked whether the participants followed the instructions properly and started the experiment. The participants read 20 texts for 2 minutes each. The total reading time, including auditory signals between the texts, was about 45 minutes. In order to enable a temporal orientation, participants received feedback on their progress in the experiment before texts 11, 16, 19, and 20. The proofreading task and the responding to the questionnaires (see Experiment 1) took about 60 minutes.

4.1.4. Design

The experiment comprised a one-factorial design with pixel density (132 vs. 264 ppi) as between-subjects variable. The dependent variables were participants' reading speed (number of words read) and proofreading performance (number of errors detected – false alarms), the ratings in the three scales of the multi-dimensional mood state questionnaire (pleasant vs. unpleasant mood, alertness vs. fatigue, calmness vs. agitation), the ratings in the three scales of the physical discomfort questionnaire (eyestrain, headache, musculoskeletal strain), and the chosen viewing distance between participants' eyes and the display.

Assuming a medium to large effect size of $d = 0.60$ for the display mode manipulation and desired levels of $\alpha = \beta = .05$, an a priori power analysis revealed that data would have to be collected from $N = 122$ participants. We were able to collect data from $N = 156$ participants so that the power was even larger than what we had planned for ($1 - \beta = .98$). All reported *t* tests for the relevant comparisons are one-sided, with the exception of the reading comfort evaluations that are based on two-tailed *t* tests.

4.2. Results

4.2.1. Proofreading performance

Independent groups *t* tests showed that there were no effects of pixel density on treading speed, $t(154) = 0.80$, $p = .21$, $d = 0.13$, and on proofreading performance, $t(154) = 0.23$, $p = .41$, $d = 0.04$ (Figure 2).

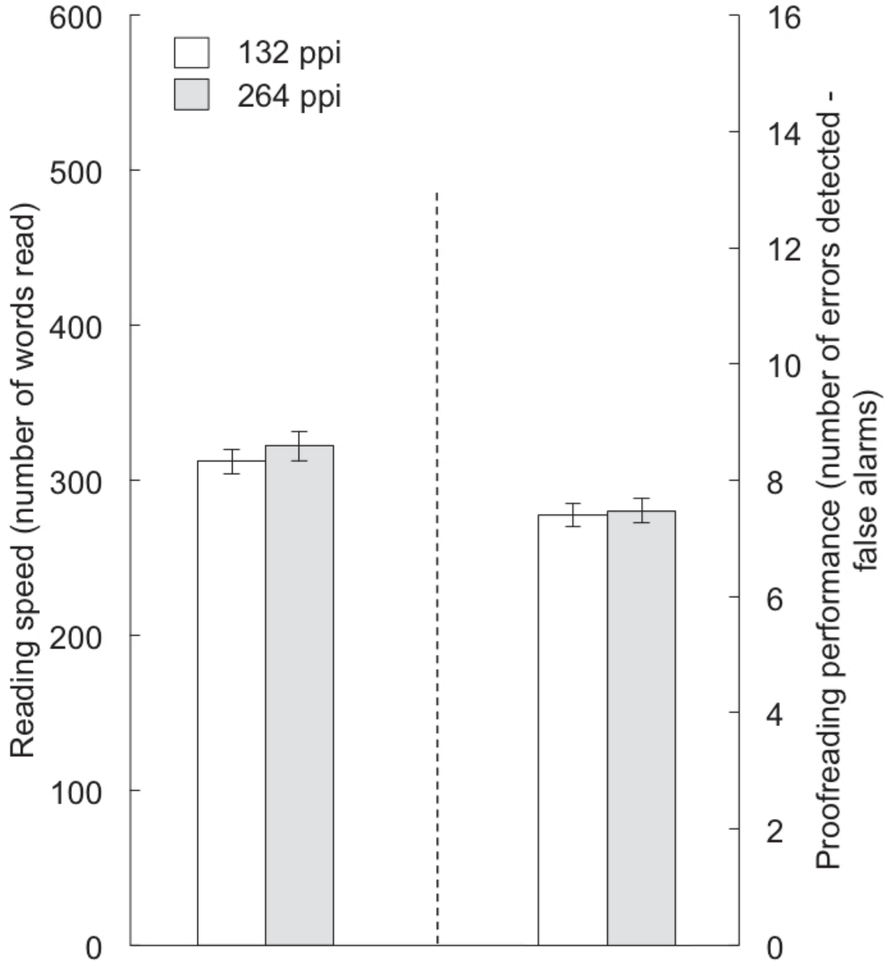


Figure 2

Mean reading speed (words read) on the left panel and mean proofreading performance (number of errors detected - false alarms) on the right as a function of pixel density for Experiment 2. Error bars represent the standard errors of the means.

4.2.2. Mood state and physical discomfort ratings

There was a descriptive trend towards a stronger increase of tiredness in the 264 ppi compared to the 132 ppi condition, but the difference did not reach the preset level of significance, $t(152) = 1.39$, $p = .08$, $d = 0.23$ (Table 2). Pixel density did not affect changes in the "pleasant versus unpleasant mood" and "calmness versus agitation" dimensions, all ts(152) < 1.22, ps > .11, ds < 0.20.

Participants in the 132 ppi condition reported a stronger increase of headache, $t(151) = 1.70$, $p = .046$, $d = 0.28$, and musculoskeletal strain, $t(151) = 1.92$, $p = .03$, $d = 0.31$, during the course of the experiment as compared to those in the 264 ppi condition (Table 3). Pixel density did not affect changes in eyestrain, $t(150) = 0.86$, $p = .20$, $d = 0.14$.

4.2.3. Viewing distance

Participants chose a similar viewing distance in the 132 ppi condition ($M = 37.93$ cm, $SE = 0.68$ cm) and in the 264 ppi condition ($M = 38.96$ cm, $SE = 0.83$ cm), $t(152) = 0.95$, $p = .17$, $d = 0.15$.

4.2.4. Evaluations of reading comfort

There were no significant differences between the two groups in evaluations of perceived contrast, $t(150) = 0.54$, $p = .59$, $d = 0.09$, brightness, $t(153) = 0.99$, $p = .32$, $d = 0.16$, comfort of reading position, $t(153) = 0.82$, $p = .41$, $d = 0.13$, glare, $t(154) = 0.41$, $p = .69$, $d = 0.07$, and reflections, $t(152) = 0.90$, $p = .37$, $d = 0.14$ (Question 3 to 7; see Table 4 for the mean values and standard errors of the means).

4.3. Discussion

The purpose of Experiment 2 was to test whether reading speed and performance would differ between displays with standard (132 ppi) and high (264 ppi) pixel densities based on the assumption that proofreading task could be more sensitive to variations in pixel densities than the text comprehension task used in Experiment 1. This was not the case. Proofreading performance and reading speed were equivalent for the standard and high-density displays, which is in line with the findings of Gujar et al. (1998), Wright et al. (1999), and Ziefle (1998, Experiment 1) and supports the assumption that pixel densities beyond a critical level no longer affect legibility in terms of visual performance.

However, there was a stronger increase in the ratings of headache and musculoskeletal strain during the course of the experiment in the 132 ppi than in the 264 ppi condition. This may indicate that participants invested more effort in the 132 ppi condition in order to maintain a certain level of proofreading speed and performance. The viewing distance and the comfort-of-reading position ratings did not differ between conditions.

5. General discussion and Conclusions

Research from the 1980s and 1990s showed that an increase in pixel density of computer displays roughly from about 60 ppi to about 130-150 ppi has positive effects on measures of visual performance, subjective well-being, and preference (as shown by Gould et al., 1987; Huang et al., 2009; Wright et al., 1999; Ziefle, 1998). For the effects of higher-resolution displays that have recently become more and more popular the literature is rather sketchy. Given this we thought it necessary to conduct a systematic investigation into the effects high-resolution displays on performance measures and on measures of subjective well-being. The results are quite clear. Reading comprehension, reading speed, and proofreading performance did not differ as a function of whether the display resolution was 132 ppi or 264 ppi. The sample sizes in the present experiments were relatively large, allowing for effects of $d = 0.6$ of the pixel density variable to be detected with a statistical power of .98 given the conventional level of $\alpha = .05$. Thus, we conclude that the absence of significant differences in visual performance is probably not the result of a Type II error.

In more general terms, then, the present results together with those reported in the literature seem to indicate that visual performance increases with pixel density roughly up to about 130-150 ppi, and then reaches an asymptote as pixel density increases further. One problem with this conclusion is that the area in which the asymptote begins is relatively indeterminate. Ideally, one would vary pixel density in small steps from about 50 ppi to about 250 ppi, Unfortunately, displays with a pixel density below 100 ppi are obsolete and, hence, no longer available (and of no practical relevance).

Participants consistently chose a viewing distance of about 40 cm, independent of the pixel density of the display from which they were reading. At this distance, the pixel sizes for the 132 ppi and 264 ppi displays are about 0.028° (192 µm) and 0.014° (96 µm), respectively. Given this, a "normal" human eye with a visual acuity of 20/20 (or 1.0) should be able to distinguish single pixels on the 132 ppi display, but not on the 264 ppi display. Thus, although the difference between both conditions must have lead to differences in the percept, this difference was not measurable in terms of reading comprehension, reading speed, and proofreading performance.

Wright et al. (1999) found differences in visual acuity between a 102 ppi CRT display, an 83 ppi TFT-LCD, and a 157 ppi TFT-LCD in a word chart that was similar to the letter charts used to measure visual acuity. These authors argued that font sizes near the legibility threshold are needed to reveal significant differences in reading speed and comprehension. This was not the case in the present experiments. Thus, one could argue that differences between standard and high pixel density displays might have been found had we used a much smaller font size. This may indeed be the case and needs to be investigated in future studies. However, it seems that using a font size that is clearly above threshold has much more ecological validity than using near-threshold font sizes given that user can normally adjust the text size on the displays of desktop computers and mobile devices.

In contrast to the reading performance measures, participants' subjective ratings of their physical discomfort provided some indication of an advantage of the high pixel density display over the standard pixel density display. There was a significantly larger increase in headache and musculoskeletal strain during the course of Experiment 2 in the standard pixel density condition as compared with the high pixel density condition. Similar trends were present at a descriptive level in Experiment 1, but these were not statistically significant. A possible explanation for the absence of significant differences in the objective measures of visual performance on the one hand and the higher subjective physical strain in the standard pixel density condition on the other hand is that participants tried to maintain a certain level of reading speed and proofreading performance which represents their “standard” acquired in many years of reading. Participants in the lower-pixel density condition may have invested more effort into the task to compensate for the less-than-perfect display, which may have resulted in more physical strain. If this speculation were valid, then fatigue should occur earlier with standard than with high pixel density displays such that a performance difference between these display types may occur after several hours of reading. Whether this is the case is another open question that needs to be investigated in subsequent studies.

In general, subjective measures may be more sensitive to legibility differences than objective measures of visual performance. For instance, Ziefle (1998) found no differences in visual performance between 60 ppi and 120 ppi display, but none of her participants preferred the 60 ppi condition. Likewise, the participants in the study by Gujar et al. (1998) showed equal proofreading performance with different pixel densities, but they rated the ease of reading from the screen with the lowest number of pixels per inch as significantly worse compared with the other conditions. Finally, Wright et al. (1999) showed stronger preferences for high than for low pixel densities, but found no differences in reading speed and comprehension. Based on these finding, it is likely that users have an overall preference for higher display resolutions. An impact of this preference on perceived display quality and, as a consequence, on visual effort or eyestrain symptoms is conceivable and should be the subject of future investigations.

In sum, the present findings support the assumption that reading performance on a screen with high pixel density (264 ppi) is as good as performance on a screens with a standard pixel density (132 ppi), at least as long as the text size is clearly above threshold and reading time does not exceed about one hour. However, certain complaints about physical strain are less likely with higher than with lower pixel densities even with reading times shorter than one hour. Thus, displays with high pixel densities (264 ppi and above) should be preferred over displays with standard (132 ppi or lower) pixel densities.

References

Alt, P. M., & Noda, K. (1998). Increasing electronic display information content: An introduction. *IBM Journal of Research and Development, 42*(3.4), 315-320. doi: 10.1147/rd.423.0315

Bridgeman, B., Lennon, M. L., & Jackenthal, A. (2003). Effects of screen size, screen resolution, and display rate on computer-based test performance. *Applied Measurement in Education, 16*(3), 191-205. doi: 10.1207/s15324818ame1603_2

Bridger, R. S. (1988). Postural adaptations to a sloping chair and work surface. *Human Factors, 30*(2), 237-247. doi: 10.1177/001872088803000210

Buchner, A., & Baumgartner, N. (2007). Text-background polarity affects performance irrespective of ambient illumination and colour contrast. *Ergonomics, 50*(7), 1036-1063. doi: 10.1080/00140130701306413

Buchner, A., Mayr, S., & Brandt, M. (2009). The advantage of positive text-background polarity is due to high display luminance. *Ergonomics, 52*(7), 882–886.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NY: Erlbaum.

Creed, A., Dennis, I., & Newstead, S. (1987). Proof-reading on VDUs. *Behaviour and Information Technology, 6*(1), 3-13. doi: 10.1080/01449298708901814

Cushman, W. H. (1986). Reading from microfiche, a VDT, and the printed page: Subjective fatigue and performance. *Human Factors, 28*(1), 63-73. doi: 10.1177/001872088602800107

Dillon, A. (1992). Reading from paper versus screens: a critical review of the empirical literature. *Ergonomics, 35*(10), 1297-1326. doi: 10.1080/00140139208967394

Eastman, M. C., & Kamon, E. (1976). Posture and subjective evaluation at flat and slanted desks. *Human Factors, 18*(1), 15-26. doi: 10.1177/001872087601800102

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191. doi: 10.3758/BF03193146

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. doi: 10.1177/001872088702900501

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. doi: 10.1177/001872088402600308

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, 42*(6), 527-531.

Heuer, H., Hollendiek, G., Kröger, H., & Römer, T. (1989). Die Ruhelage der Augen und ihr Einfluß auf Beobachtungsabstand und visuelle Ermüdung bei der Bildschirmarbeit. [Resting position of the eyes and its influence on observation distance and visual fatigue during viusal display work]. *Zeitschrift für experimentelle und angewandte Psychologie, 36*(4), 538-566.

Holzinger, A., Baernthaler, M., Pammer, W., Katz, H., Bjelic-Radisic, V., & Ziefle, M. (2011). Investigating paper vs. screen in real-life hospital workflows: Performance contradicts perceived superiority of paper in the user experience. *International*

Journal of Human-Computer Studies, 69(9), 563-570. doi: 10.1016/J.Ijhcs.2011.05.002

Huang, D.-L., Rau, P.-L., & Liu, Y. (2009). Effects of font size, display resolution and task type on reading Chinese fonts from mobile devices. *International Journal of Industrial Ergonomics, 39*(1), 81-89. doi: 10.1016/j.ergon.2008.09.004

Köpper, M., Mayr, S., & Buchner, A. (2015). *Reading from computer screen versus reading from paper: Does it still make a difference?* Manuscript submitted for publication.

Kretzschmar, F., Pleimling, D., Hosemann, J., Füssel, S., Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2013). Subjective impressions do not mirror online reading effort: concurrent EEG-eyetracking evidence from the reading of books and digital media. *PLoS ONE, 8*(2), e56178. doi: 10.1371/journal.pone.0056178

Margolin, S. J., Driscoll, C., Toland, M. J., & Kegler, J. L. (2013). E-readers, computer screens, or paper: Does reading comprehension change across media platforms? *Applied Cognitive Psychology, 27*(4), 512-519. doi: 10.1108/02640471211241663

Mayr, S., & Buchner, A. (2010). After-effects of TFT-LCD display polarity and display colour on the detection of low-contrast objects. *Ergonomics, 53*(7), 914–925.

Menozzi, M., Lang, F., Näpflin, U., Zeller, C., & Krueger, H. (2001). CRT versus LCD: effects of refresh rate, display technology and background luminance in visual performance. *Displays, 22*(3), 79-85. doi: 10.1016/S0141-9382(01)00054-3

Menozzi, M., Näpflin, U., & Krueger, H. (1999). CRT versus LCD: A pilot study on visual performance and suitability of two display technologies for use in office work. *Displays, 20*(1), 3-10. doi: 10.1016/S0141-9382(98)00051-1

Meyer, B. J. F., & Poon, L. W. (1997). Age differences in efficiency of reading comprehension from printed versus computer-displayed text. *Educational Gerontology, 23*(8), 789-807.

Mills, C. B., & Weldon, L. J. (1987). Reading Text from Computer Screens. *Computing Surveys, 19*(4), 329-358.

Moore, D., & Zabrucky, K. (1995). Adult Age-Differences in Comprehension and Memory for Computer-Displayed and Printed Text. *Educational Gerontology, 21*(2), 139-150.

Näsänen, R., Karlsson, J., & Ojanpää, H. (2001). Display quality and the speed of visual letter search. *Displays, 22*(4), 107-113. doi: 10.1016/S0141-9382(01)00058-0

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

Piepenbrock, C., Mayr, S., & Buchner, A. (2014b). Smaller pupil size and better proofreading performance with positive than with negative polarity displays. *Ergonomics, 57*(11), 1670-1677. doi: 10.1080/00140139.2014.948496

Piepenbrock, C., Mayr, S., Mund, I., & Buchner, A. (2013). Positive display polarity is advantageous for both younger and older adults. *Ergonomics, 56*(7), 1116-1124. doi: 10.1080/00140139.2013.790485

Shieh, K.-K., & Lin, C.-C. (2000). Effects of screen type, ambient illumination, and color combination on VDT visual performance and subjective preference. *International Journal of Industrial Ergonomics, 26*(5), 527-536. doi: 10.1016/S0169-8141(00)00025-1

Steyer, R., Schwenkmezger, P., Notz, P., & Eid, M. (1997). *Der mehrdimensionale Befindlichkeitsfragebogen (MDBF) [The multidimensional mood questionnaire (MDBF)].* Göttingen: Hogrefe.

Wilkinson, R. T., & Robinshaw, H. M. (1987). Proof-reading: VDU and paper text compared for speed, accuracy and fatigue. *Behaviour and Information Technology, 6*(2), 125-133. doi: 10.1080/01449298708901822

Wright, S. L., Bailey, I. L., Tuan, K. M., & Wacker, R. T. (1999). Resolution and legibility: A comparison of TFT-LCDs and CRTs. *Journal of the Society for Information Display, 7*(4), 253-256.

Ziefle, M. (1998). Effects of Display Resolution on Visual Performance. *Human Factors, 40*(4), 554-568. doi: 10.1518/001872098779649355

Ziefle, M. (2001). Aging, visual performance and eyestrain in different screen technologies. *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting*, 262-266. doi: 10.1177/154193120104500315

Eidesstattliche Erklärung

Hiermit versichere ich an Eides Statt, dass ich die Arbeit mit dem Titel „ Lesen am Bildschirm - Ein systematischer Vergleich von TFT-LCDs mit Papier und der Einfluss hoher Bildschirmauflösungen auf objektive Leistungsmaße und subjektives Wohlbefinden“ selbstständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf“ verfasst habe. Ich versichere insbesondere:

(1) Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt.

(2) Alle wörtlich oder dem Sinn nach aus anderen Texten entnommenen Stellen habe ich als solche kenntlich gemacht; dies gilt für gedruckte Texte ebenso wie für elektronische Ressourcen.

(3) Die Arbeit habe ich in der vorliegenden oder einer modifizierten Form noch nicht als Dissertation vorgelegt – sei es an der Heinrich-Heine-Universität Düsseldorf oder an einer anderen Universität.

Datum: 31. August 2015

Name: Maja Köpper

Unterschrift: