

# Besser Lesen am Bildschirm?

Nutzerbeobachtung mittels Eyetracking und Interviews geben Einblicke in das Online-Lesen

Birgit Fuhrmann

ZHAW Zürcher Hochschule für Angewandte

Wissenschaften

Zürich, Schweiz

birgit.fuhrmann2@zhaw.ch

Martin Schuler

ZHAW Zürcher Hochschule für Angewandte

Wissenschaften

Zürich, Schweiz

martin.schuler@zhaw.ch

## ZUSAMMENFASSUNG

Ziel der transdisziplinären Vorstudie war es, herauszufinden, ob ein Softwareprodukt mithilfe von spezieller Textauszeichnung die Informationsverarbeitung beim Online-Lesen verändern kann. Mittels Nutzerbeobachtung wurden quantitative Daten zur Lesegeschwindigkeit und Behaltensleistung und qualitative Daten zum Leseverhalten erhoben. Die Ergebnisse zeigten eine niedrige Lesegeschwindigkeit und keine signifikanten Unterschiede zwischen ausgezeichneten und normalen Texten. Gleichzeitig zeigte sich bei der Behaltensleistung eine messbare Differenz und die Eyetracking-Daten lieferten einen Hinweis für eine veränderte Informationsaufnahme. Die Interviews gaben einen Einblick in das individuelle Leseverhalten. Kurzum, die Vorstudie konnte erste Anhaltspunkte aufzeigen, dass das Softwareprodukt die Informationsverarbeitung verändern kann. Weiterhin lieferte sie wertvolle Erkenntnisse für das Testdesign künftiger Studien, die auch für PraktikerInnen in der Produktentwicklung interessant sind.

## KEYWORDS

Online-Lesen, Nutzerbeobachtung, Eyetracking, Behaltensleistung, Leseverhalten

## 1 PROJEKTZIEL UND AUSGANGSLAGE

Bei dieser Vorstudie handelt es sich um die Evaluation eines Softwareprodukts, das Leserinnen und Leser beim Lesen und Verarbeiten von digitalen Texten unterstützen und lenken soll. Mithilfe des Softwareprodukts soll die Lesbarkeit von Online-Texten erhöht werden, d. h. dadurch soll es Lesenden ermöglichen, digitale Texte effizienter zu lesen. Als Methodik wurde die Nutzerbeobachtung in Anlehnung an das Usability-Testing gemäss [1] ausgewählt. Es handelt sich hierbei um eine empirische Methode, die mithilfe von induktivem Testing eine formative Evaluation erlaubt. Der Untersuchungsgegenstand waren mit der Software überarbeitete Textsamples. Die Nutzerbeobachtung wurde zusätzlich mithilfe des Einsatzes von Eyetracking durchgeführt, eine methodische Herangehensweise innerhalb des Usability-Testings, die die Blickbewegungen der Testpersonen aufzeichnet.

Ziel der Zusammenarbeit war es, im Usability-Labor mittels Nutzerbeobachtung herauszufinden, ob das Softwareprodukt einen messbaren Mehrwert in Bezug auf die Informationsverarbeitung

bietet und mit welchen messbaren Kriterien dies nachgewiesen werden kann.

### 1.1 Testpersonen

Als Anforderungen an die Testpersonen wurde gestellt, dass sie regelmässig wissenschaftliche Texte am Bildschirm lesen. Daher bestand die ausgewählte Testgruppe aus Studierenden des Bachelorstudiengangs Angewandte Sprachen an der ZHAW. Die Auswahl erfolgte zusätzlich mit dem Ziel, eine möglichst homogene Zusammenstellung zu gewährleisten. Einerseits, um die Ergebnisse der Testpersonen untereinander vergleichen zu können und andererseits, da sie der primären Zielgruppe des Softwareprodukts am besten entsprachen: digital Natives, die viel Zeit mit Online-Lesen von (wissenschaftlichen) Texten verbringen und Erfahrungen im Umgang mit digitalen Ausgabegeräten mitbringen. So konnten u.a. Einflussfaktoren auf den Leseprozess z. B. durch technische Hemmnisse ausgeschlossen werden. Basierend auf der Einschätzung der Testleitung teilte sich die Testgruppe in 11 weibliche und 1 männliche Testpersonen auf, das Geschlecht wurde nicht spezifisch abgefragt, da es für die Vorstudie nicht relevant war. Im Sinne der Diversität soll angemerkt werden, dass divers als Geschlecht eingeschlossen sein könnte.

### 1.2 Messkriterien

Für den quantitativen Teil der Untersuchung wurden 3 messbare Kriterien bestimmt: Lesegeschwindigkeit, Textverständnis bzw. Behaltensleistung und Regressionen. Dies anhand von zwei inhaltlich unterschiedlichen Texten, die sich in der Textkomplexität sehr ähnlich waren. Die Textkomplexität wurde mit dem Lesbarkeitsindex gemäss [2] bestimmt. Die Lesegeschwindigkeit wurde mittels folgender Formel berechnet und erhoben: Lesegeschwindigkeit = Textlänge/Lesezeit (Wörter/Minute).

Das Textverständnis bzw. Behaltensleistung wurde anhand von Fragen zum Text analysiert, dabei wurde bei der Auswertung nicht zwischen Verständnis und Behaltensleistung unterschieden, da es für die Vorstudie nicht relevant erschien. Die Antworten auf die Textfragen wurden anschliessend ausgewertet mit einer maximal zu erreichenden Punktzahl von 40 Punkten. Mithilfe von Eyetracking wurde die Anzahl der Regressionen erhoben. Regressionen sind Rücksprünge im Text, hier wurden nur Regressionen über eine oder mehrere Zeilen hinweg beachtet, nicht solche innerhalb einer Zeile.

Eine besondere Herausforderung lag im Nutzungskontext: Die Auswahl der beiden unterschiedlichen Ausgabegeräte für die Leseraufgabe – Desktop-PC und Tablet – basiert auf einem Konsens mit dem Projektpartner für die Interessenswahrung aller Projektbeteiligten und Zielsetzung der Vorstudie. Einerseits entspricht

das Lesen am Tablet ohne Eyetracking-Aufzeichnung eher dem gewöhnlichen Nutzungskontext, andererseits war Eyetracking als die zentrale Erhebungsmethodik für die Vorstudie festgelegt worden.

## 2 VORGEHEN

### 2.1 Testsetting

Die Tests fanden im Januar und Februar 2022 im Usability-Labor der ZHAW in Winterthur statt und es wurde folgende Beobachtungs- und Evaluationsinstrumente genutzt (s. [Tabelle 1](#)).

**Tabelle 1: Beobachtungs- und Evaluationsinstrumente**

Ausgabegeräte für Testpersonen	Desktop-PC/Tablet
Eyetracking-Aufzeichnung	Stationärer Eyetracker am PC
SMI	ExperimentCenter
Videoaufzeichnung	Domekameras
Audioaufzeichnung	Stationäres Mikrofon
Screen-Recording	OBS Studio
Interviews	Fragebogen

### 2.2 Stimuli

Die Kriterien für die Textauswahl wurden wie folgt definiert:

- (1) Populärwissenschaftlich
- (2) Monothematisch
- (3) Kein Aktualitätsbezug
- (4) Vergleichbare Komplexität

Die folgenden beiden Texte aus den Online-Archiven von «NZZ Folio» und «Die Zeit» erfüllen alle diese Anforderungen und wurden daher für diesen Test ausgewählt:

- (1) Text «Schnabeltier»: Wenn Säugetiere Eier legen. NZZ Folio Juli 2008. [3]
- (2) Text «Labormäuse»: Schöner wohnen für Labortiere. DIE ZEIT 47/2009. [4]

Gemäss dem von [2] postulierten Lesbarkeitsindex weisen beide Texte eine vergleichbare Komplexität auf:

- (1) 37 (Schnabeltier) zu
- (2) 35 (Labormäuse).

Beide Texte wurden für das Lesen am Bildschirm angepasst (Schriftart und -grösse, Zeilenabstand, Zeilenlänge). Von beiden Texten wurde eine Variante mit Textauszeichnung (A-Text) sowie eine normale Variante (N-Text) erstellt. Abgesehen von dem Unterscheidungsmerkmal der Textauszeichnung war das Layout aller Textvarianten identisch.

### 2.3 Testablauf

Die Testpersonen hatten den Auftrag, zwei Texte zu lesen und jeweils im Anschluss daran inhaltliche Fragen zum Gelesenen mündlich zu beantworten. Ziel der inhaltlichen Fragen war die Erhebung der Textverständnis bzw. Behaltensleistung der Testpersonen zu den beiden gelesenen Texten. Einen der Texte lasen sie auf einem Tablet, den anderen auf einem Desktop-PC.

Zur Minimierung von Gewöhnungseffekten und anderen aus dem Setting resultierenden Störeffekten wurden alle Parameter randomisiert und gleichmässig auf alle Testpersonen verteilt. Für

die Lektüre der beiden Texte waren jeweils 10 Minuten vorgesehen, die Testpersonen wurden nicht unterbrochen, wenn sie diese Dauer überschritten. Der Test war auf eine Dauer bis max. 90 Minuten angelegt.

### 2.4 Pre- und Post-Task-Interview

Für den qualitativen Teil der Untersuchung wurde vor Beginn der Lese-Aufgaben eine kurze Befragung der Testpersonen durchgeführt, um aufgrund ihrer Selbsteinschätzung Informationen über ihr individuelles Leseverhalten zu erfahren. Im Anschluss an die beiden Lese-Aufgaben wurden die Testpersonen zu ihren individuellen Eindrücken und Erfahrungen befragt, um Erkenntnisse zur Einschätzung des Leseverhaltens und dem Umgang mit Lesen am Bildschirm gewinnen zu können. Die Testpersonen wiesen laut Angaben keine Einschränkungen beim Lesen auf.

## 3 DISKUSSION DER ERGEBNISSE

Die messbaren Erkenntnisse, die aus der Projektphase des Testings gewonnen werden konnten, weisen lediglich marginale Unterschiede zwischen Texten mit und ohne Textauszeichnung auf. Gleichwohl hat diese Vorstudie einige Tendenzen aufzeigen können, die auf Auswirkungen des Softwareprodukts auf den Leseprozess hinweisen.

### 3.1 Lesegeschwindigkeit

Hinsichtlich der Lesegeschwindigkeit konnten die Testpersonen die A-Texte lediglich marginal schneller lesen als die N-Texte (s. [Abbildung 2](#)). Um den Einfluss anderer Parameter auf die Lesegeschwindigkeit auszuschliessen, wurden die Durchschnittswerte zusätzlich unter den folgenden Gesichtspunkten betrachtet:

- (1) A-Text versus N-Text
- (2) Text 1 versus Text 2
- (3) Desktop-PC versus Tablet
- (4) Task 1 versus Task 2
- (5) Mit versus ohne TP10

Auch unter diesen Perspektiven lässt sich kein signifikanter Unterschied der durchschnittlichen Lesegeschwindigkeiten feststellen.

Die durchschnittliche Lesegeschwindigkeit über alle Testpersonen liegt bei 95-192 Wörtern pro Minute, was eine verhältnismässig niedrige Geschwindigkeit ist in Bezug auf die Lesegeschwindigkeit von Studierenden. Diese liegt gemäss [5] bei 150-400 Wörtern pro Minute. Eine mögliche Erklärung für diese vom Normal abweichende Lesegeschwindigkeit könnte sein, dass die Testpersonen aufgrund der Testsituation unter Zeit- und Leistungsdruck (Lesen und Beantwortung von Fragen) und unter Beobachtung langsamer und genauer versuchten zu lesen als in ihrem gewöhnlichen Nutzungskontext.

Besonders interessant war die Lesegeschwindigkeit einer Testperson (TP10), die stark von den anderen und des Durchschnittswerts abwich (s. [Abbildung 1](#)). Die Lesegeschwindigkeit sticht zwar heraus, dennoch konnte beim Textverständnis bzw. Behaltensleistung der Testperson kein wesentlicher Unterschied festgestellt werden: die Werte liegen hier knapp unter dem Durchschnittswert, d.h. das Textverständnis bzw. Behaltensleistung war weder höher noch tiefer. Es sei darauf hingewiesen, dass die Auswertung der Lesegeschwindigkeit daher sowohl unter Einbezug der Werte von TP10 als

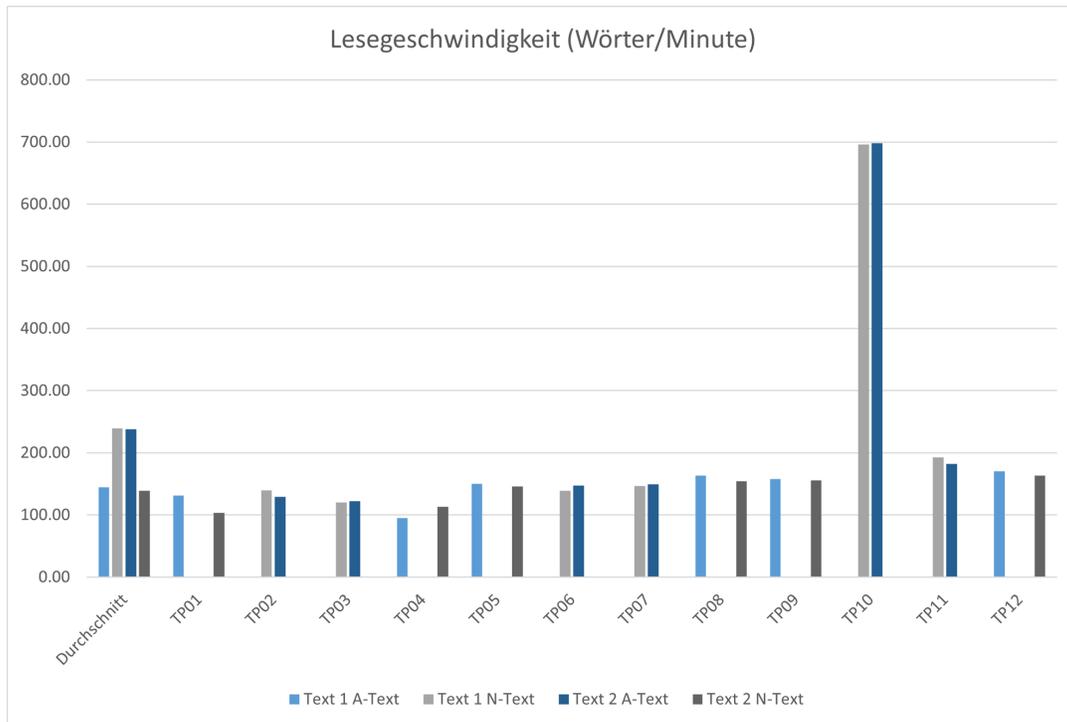


Abbildung 1: Lesegeschwindigkeit Durchschnittswert pro Textsample und Einzelwerte pro Testperson (A-Text/N-Text)

auch unter Abzug dessen erfolgte, um eine mögliche Beeinflussung ausschliessen zu können.

### 3.2 Textverständnis/Behaltensleistung

Mithilfe des gemessenen Textverständnisses bzw. Behaltensleistung konnte gezeigt werden, dass es beim Inhalt der A-Texte etwas höher ist (s. Tabelle 2).

Tabelle 2: Testverständnis/Behaltensleistung in Durchschnittswerten pro Text und für beide Texte

Auszeichnung	Durchschnitt		
	Text 1	Text 2	Text 1 & 2
A-Text	12.00	11.50	11.75
N-Text	9.00	10.67	9.83

### 3.3 Regressionen

Die Analyse der Eyetracking-Daten erfolgte unter Vorbehalt, da keine 2 Datensamples pro Testperson vorlagen, sondern jeweils ein Sample entweder zum A- oder N-Text, was einen individuellen Vergleich pro Testperson unmöglich machte. Es gab keine deutlichen Unterschiede zwischen den beiden Text-Varianten bezüglich Anzahl Regressionen: hier liegen die Durchschnittswerte bei beiden Varianten sehr nahe beieinander (s. Tabelle 3).

Tabelle 3: Vergleich durchschnittliche Regressionen pro Text

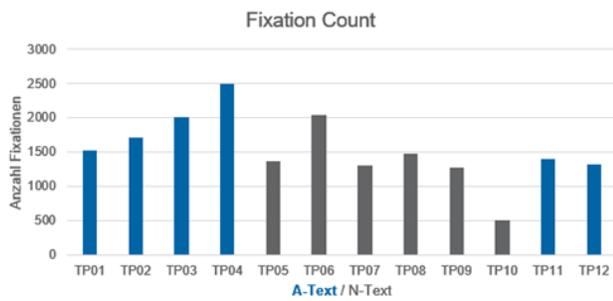
Auszeichnung	Durchschnitt Text
A-Text	33.00
N-Text	33.80

### 3.4 Fixationen

Die Werte zur Anzahl der Fixationen der einzelnen Testpersonen zeigt sehr deutlich, wie individuell die Blickbewegungen sind (s. ??): die Werte schwanken von sehr hoch mit ca. 2500 (TP04) bis niedrig von 500 (TP10). Bei den A-Texten liegt die durchschnittliche Anzahl Fixationen deutlich über dem Wert der N-Texte (s. Tabelle 4), wohingegen die durchschnittliche Fixationsdauer tiefer liegt (s. Tabelle 5). Das könnte ein erster Hinweis darauf sein, dass beim Lesen mit dem Softwareprodukt mehr Informationen aufgenommen werden können (mehr Fixationen in gleicher Zeit) und gleichzeitig die Informationsaufnahme schneller abläuft (kürzere Dauer der Fixationen). Jedoch ist die Aussagekraft der Eyetracking-Daten aufgrund des Testsettings unter Vorbehalt zu bewerten: dies könnte eine rein zufällige Verteilung sein, die auf den individuellen Blickverläufen der Testpersonen beruht.

### 3.5 Wertvolle Erkenntnisse zum Testdesign

Aufgrund der Durchführung der Lese-Aufgaben mit zwei verschiedenen Ausgabegeräten – Desktop-PC und Tablet – konnten wertvolle Erfahrungen zum Testdesign für künftige Projekte in diesem Bereich gesammelt werden. Der Einsatz der beiden Ausgabegeräte



**Abbildung 2: Anzahl Fixationen pro Testperson bezüglich Task, die am Desktop-PC durchgeführt wurde.**

**Tabelle 4: Durchschnittliche Anzahl Fixationen über alle Testpersonen (ohne TP10)**

Auszeichnung	Durchschnitt Text
A-Text	1740.67
N-Text	1489.00

**Tabelle 5: : Durchschnittliche Fixationsdauer über alle Testpersonen (ohne TP10)**

Auszeichnung	Durchschnitt Text
A-Text	281.62
N-Text	296.64

erwies sich jedoch in Bezug auf die Auswertung der Eyetracking-Daten als nicht optimal, da pro Testperson jeweils nur eine der beiden Lese-Aufgaben mit Eyetracking aufgezeichnet werden konnte. Dadurch war es nicht möglich, einen direkten Vergleich des Leseverhaltens bei A- und N-Text pro Testperson auszuwerten. Dies ist insofern essenziell, als dass die Blickbewegungen bei den Testpersonen sehr individuell sind, so dass es den direkten Vergleich pro Testperson für die Auswertung benötigt, um daraus dann ggf. in einem zweiten Schritt Aussagen über alle Testpersonen ableiten zu können. Für zukünftige Projekte muss hier die Empfehlung gegeben werden, die Eyetracking-Aufzeichnung über die Gesamtheit der

Test-Samples durchzuführen, um eine vollumfängliche Auswertung der Eyetracking-Daten und aussagekräftige Erkenntnisse daraus zu ermöglichen.

### 3.6 Reflexion und Ausblick

Bei der vorliegenden Vorstudie sind aufgrund deren begrenztem Projektumfang und der kurzen Umsetzungszeit innert 3 Monaten einige Einschränkungen vorhanden. Die Begrenzung auf eine bestimmte Zielgruppe (Homogenität) und die geringe Anzahl der Testpersonen schränken die Übertragbarkeit der hier gewonnenen Erkenntnisse auf andere Zielgruppen oder breitere Bevölkerungsschichten ein. Gleichzeitig schmälert dies nicht die Erkenntnisse, denn die Vorstudie hat ihr Ziel erfüllt, einige Tendenzen aufzuzeigen. Diese könnten in grösser angelegten Projekten oder Langzeit-Untersuchungen detaillierterer untersucht werden. Darüber hinaus zeichnen sich Untersuchungen mit spezifischen Zielgruppen, beispielsweise Personen mit Dyslexie, als besonders vielversprechend ab.

### Danksagung

Bei dieser Vorstudie handelt es sich um ein anwendungsorientiertes Forschungsprojekt mit einem Praxispartner, das mit Unterstützung und Förderung der Innosuisse – schweizerischen Agentur für Innovationsförderung umgesetzt wurde.

### LITERATUR

- [1] Florian Sarodnick und Henning Brau. 2011. Methoden der Usability Evaluation. Wissenschaftliche Grundlagen und praktische Anwendung. 2. Überarbeitete und aktualisierte Auflage. Verlag Hans Huber, Bern.
- [2] Rudolf Flesch. 1948. A New Readability Yardstick. In: Journal of Applied Psychology. 32, Nr. 3, 1948, S. 221–233.
- [3] Herbert Cerutti. 2008. Wenn Säugetiere Eier legen. NZZ Folio Juli 2008. Zitiert 09.03.2022 von <https://www.nzz.ch/folio/wenn-sauegetiere-eier-legen-ld.1623293>
- [4] Magdalena Hamm. 2009. Schöner wohnen für Labortiere. DIE ZEIT 47/2009. Zitiert 09.03.2022 von <https://www.zeit.de/2009/47/N-Labortiere>
- [5] Walsch, Dominique. 2005. Beschreibung des Konstruktes 'Lesekompetenz'. Studienarbeit. Reihe: Geisteswissenschaften. Verlag GRIN, München.