Héctor Mangas Afonso (2389916)
Utrecht University

Improving paragraph reading:
The effect of beginning-of-line additions in readability, an eye tracking experiment

Submitted July 25, 2022
(27.5 ECTS)

Supervised by
Jessica Heeman

Auditored by
Krista Overvliet

Improving paragraph reading:
The effect of beginning-of-line additions in readability, an eye tracking experiment

Abstract

From a perceptual point of view, a paragraph is a very complex and tangled unit of information. Paragraphs contain a multitude of shapes that can't be interpreted all at once. Precise navigation is thus essential. The present research tests three conditions expected to affect navigation within paragraphs, both based on previous literature and editorial design history. In the 15th century, a common practice was to add the first word of the next page at the bottom of the current one (De Hamel, 1992). Contemporary research has shown within a block of text, line differentiation greatly affects its readability, as return sweeps (the eye-movement to transition between lines) can be performed more precisely (Bhatia et al., 2014; Rello et al., 2016). The present study therefore looks at how additional elements at the beginning of each line affect readability. Following intuitive knowledge from the typesetting community, the repeating of words was considered. Two design choices - one embedded and one outside were used. Another line differentiation condition, which involved a graphic addition (diamonds before each line) was also studied. It was expected to observe an improvement in reading performance against a control condition. Results showed considerable reduction in reading time, amount of fixations and leftward saccades for the diamond condition against control. The two text addition conditions, though, performed worse than control. These results show word repetition is not useful, as it seems to add cognitive load. Meanwhile, simple and distinct shapes before lines do help. Conclusions align with previous research, which suggests back sweep performance is linked with parafoveal information (Slattery & Parker, 2016), visual area in which reading acuity is at a lowest (Sandberg et al., 2008).

Research context and background

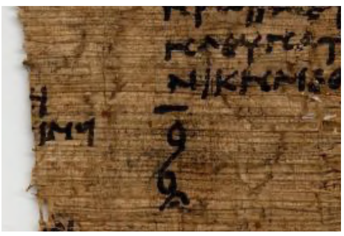


Figure 1. Close up of a copy of Menander's *Sicyonians* (3th century BC). The horizontal line, paragraphos, indicates a change in speaker. Image from Jean Gascou of the Institut de Papyrologie, Paris Sorbonne.

A paragraph is a block of text with one or more sentences. Paragraphs are usually written in prose and are divided by breaks, which physically separate them and help the reader understand the conceptual flow of the text (*Paragraphs*, 2018). But the concept of paragraph we know today has widely changed across ages. The origin of using a physical mark to divide ideas has been around since Ancient Greece (Enos, 2006). The word paragraph, indeed, comes from this time (*para-,* "beside" and *graphos,* "write"). Greek texts from around the fourth century BC show a horizontal line in the margin to the left of the main text (see Figure 1). This line, called *paragraphos,* normally meant a change of topic or structure in the text, but its meaning varied with context and depended on the scribe and author's criteria (Lunsford, 2008). This seems to be the first known effort to ease the flow of reading.

Romans took the *paragraphos* and evolved it to *kaput* ("head" in Latin), which was used to divide text in a similar fashion (Parkes, 1993). *Kaputs* were represented with the letter "K". To divide larger sections, the Latin word *capitulum* ("little head") was used, represented by the letter "C". In the middle ages, "C" completely replaced "K" as the symbol for a new block of text. Typesetters then added two vertical bars to stylise it, which evolved into the symbol we use today, the modern pilcrow (¶; Lunsford, 2008). Before the 14th century, scribes and rubricators would work together to produce manuscripts (Lewis, 1894). Many steps were used when producing these very decorated written pieces. Scribes would normally complete the body text in black ink and then rubricators added adornments and capitals in other colours. Scribes would leave blank spaces for rubricators to draw the pilcrow between blocks. With the introduction of the printing press, the process sped up and rubricators no longer had time to add pilcrows to the text, prioritising the adornment of capitals and other elements. This left spaces between blocks of text blank. The paragraph we know today was born, almost as a by-product of the printing press. From a contemporary perspective, the paragraph is essential to convey complex ideas. Readers of today expect units of information to be grouped in different paragraphs, which piece together the greater and more intricate message the whole text aims to transmit (Whitford, Titone, 2014).

But perceptually, paragraphs are still very complex blocks of information. A paragraph contains such a large amount of shapes it's impossible to interpret them all at once (Castello, 2008). For this reason, sufficiently good navigation is needed for comprehension. The present research tries to shed light on this topic by first looking at historical typesetting methods to improve navigation, both between and within the paragraph, and then testing these tools together with contemporary knowledge about paragraph reading.

Designers from different times have found ways to improve navigation within the paragraph. Newspaper typesetters, for example, introduced columns to ease reading during the 19th century. The newspaper column, in comparison to page-width text, helps navigation by reducing the distance between the end of each line and the beginning of the next (*Archive*, 2012). In more recent times, line differentiation has consistently been observed to help readability (Bhatia et al., 2014; Rello et al., 2016). Spacing between lines, for example, which makes them perceptually more salient, has a direct correlation with readability (Bhatia et al., 2014; Rello et al., 2016). Additionally, justifying text has been observed to negatively affect readability, as it decreases differences between lines (Rello et al., 2016).

The contemporary market has great interest in improving efficiency of reading, which can be observed in the incessant flow of new typographic ideas and technologies regularly coming from the type community. Fonts like the Focus Sans or Dyslexie aim to help individuals by introducing increased proportions for character differentiation and other tweaks for their intended target (these examples are aimed at individuals with ADHD and dyslexia respectively).

Figure 2. How text looks when using BeeLine Reader, a tool aimed at improving readability (Obtained from beelinereader.com).

New reading tools like BeeLine Reader are also constantly on the agenda. This one, for example, comes in the form of a paid Chrome extension and adds coloured gradients to sentences, which are aimed to "guide the reading pattern" (see Figure 2). Colour of the gradient and other variables can be changed. BeeLine Reader claims to be backed by three

This **is** **h**ow **Bio**nic **Rea**ding **te**xt **lo**oks.

Figure 3. How text looks when using Bionic Reading, a tool aimed at improving readability (Obtained from bionic-reading.com).

(unpublished) studies by three American universities, which all allegedly showed some improvement in reading. Another commercial project that has recently gained a lot of attention is Swiss designer Renato Casutt's "Bionic Reading". It consists of highlighting the first few letters of each word to "guide the eyes through a text" (see Figure 3). This software comes as an API and paid extension, and similarly to BeeLine Reader, allows to tweak a series of variables. As of now and according to Bionic Reading's website, only one unpublished and inconclusive study by an unreferenced Swiss university has been performed on Bionic reading. BeeLine Reader and Bionic Reading both were created by people in the creative sector, with very poor research later added to both projects ad hoc. There's no mention to pilots or literature research in the making of neither of these products. Both projects, which label themselves as "cognitive boosters" and "science-backed" products, lack rigorous research.

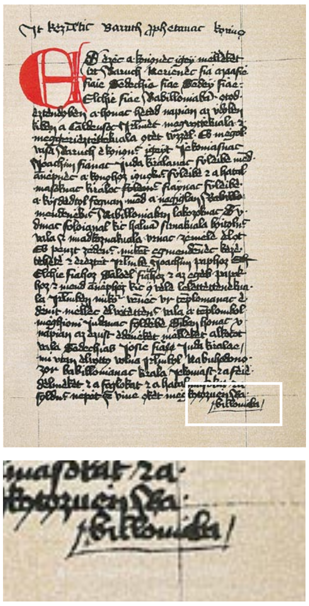


Figure 4. Fragment from a manuscript of The Death Speech, Hungarian poem from the Pray Codex, dated between the 12th and 13th century. The last word, salient from the block of text, is the first one from the next page.

When looking at how workers of type have tackled this issue across history, a common resource can be found: the repetition or cueing of words. The first instances of this come from as early as the first century, when scribes would add the first word or sentence of the next papyrus roll at the end of the current one (West, 1963). This way, the reader could quickly know the order in which the papyruses were intended to be read. This was called *reclamans,* and it was intermittently used until the 14th century, when the movable printing press was introduced (West, 1963). An example of *reclamans* can be seen in Figure 4.

With the printing press, the use of *reclamans* was lost. A stylistic shift happened, and character setting was now the norm. It wasn't until the late fifteenth century, when catchwords were reintroduced, now with a new added purpose (see Figure 5). *Reclamans* were added again so that bookbinders could easily know if they were bounding the pages in the right order (De Hamel, 1992). This practice became widespread in the mid sixteenth century, and remained popular until the industrialisation of printing, in the late eighteenth century. It's also worth mentioning here page numbering, another tool to help book setters.

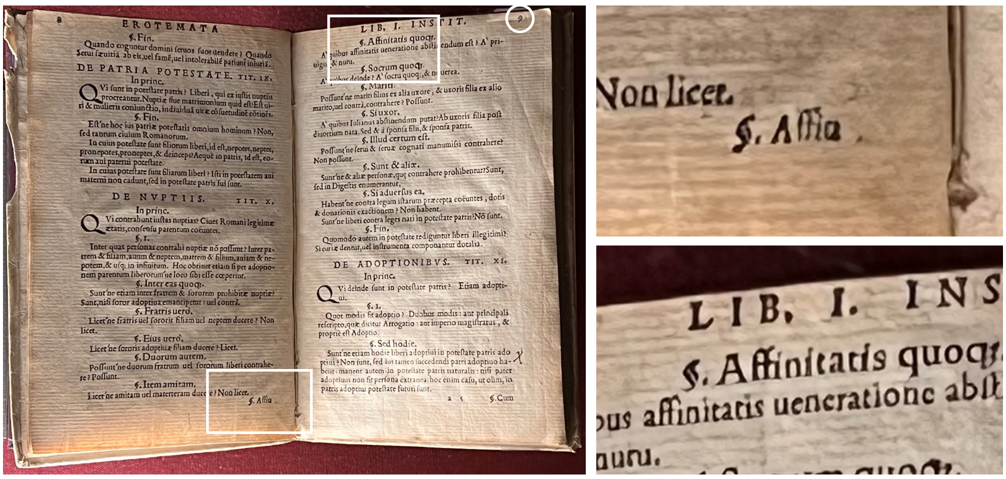
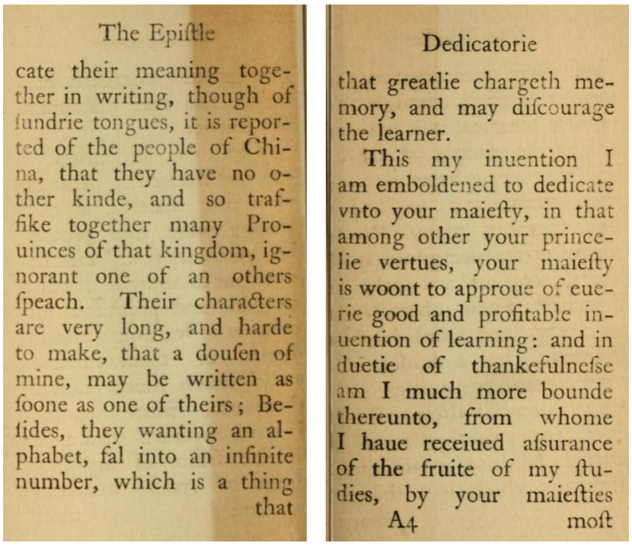


Figure 5. Two pages from Erotemata (Sébastien Gryphe, 1544), displayed at the Library of Lyon. Catchwords can be observed at the end of each of the pages (rectangles). See how this book also has its pages numbered (circle). Image courtesy of Peter Bilak.

Pages were numbered way before the appearance of catchwords (De Vinne, 1901), around the eight century, coexisted with *reclamans* and catchwords and outlived them until our days.


The Epiſtle

cate their meaning toge-
ther in writing, though of
ſundrie tongues, it is repor-
ted of the people of Chi-
na, that they have no o-
ther kinde, and so traf-
fike together many Pro-
uinces of that kingdom, ig-
norant one of an others
ſpeach. Their characters
are very long, and harde
to make, that a douſen of
mine, may be written as
ſoone as one of theirs; Be-
ſides, they wanting an al-
phabet, fal into an infinite
number, which is a thing
that

Dedicatorie

that greatlie chargeth me-
mory, and may diſcourage
the learner.
This my inuention I
am emboldened to dedicate
vnto your maieſty, in that
among other your prince-
lie vertues, your maieſty
is woont to approue of eue-
rie good and profitable in-
uention of learning: and in
duetie of thankefulneſse
am I much more bounde
thereunto, from whome
I haue receiued aſsurance
of the fruite of my ſtu-
dies, by your maieſties
A4 moſt

Figure 6. Another example of catchwords from a posterior book, Characterie, originally published in 1888. Obtained from archive.org.

After the industrialisation of printing, catchwords were secluded to more alternative and smaller publications (see Figure 6). They remained in some use until the end of the 18th century. Contemporary research has looked at the way readers perceive these repeated elements in a text, and whether they hinder readability. One of these repetitions is such of words in different lines. Using a word at the end of a sentence and repeating it at the beginning of the following is called anadiplosis (see Figure 7), and it's long been used in literary texts as a stylistic device. Generally, readers seem to ignore repetitions if the words are coherent with the text (Ehrlich and Rayner, 1981; Drieghe, et al., 2010; Rayner, et al., 2011). Another variable that affects skipping rates is character length. Longer and contextually more salient words are skipped less and seem to hinder reading (Veldre, et al., 2020). These repetitions don't seem to hinder readability but no research has looked at whether they help reading.

I know what **I like**
I like what I know.

Figure 7. An example of anadiplosis, where the last element of a line is repeated at the beginning of the next.

When it comes to lay out, literature also seems to agree on what the ideal parameters are. On one side, there seems to be a positive correlation between line spacing and readability (Nanavati and Bias, 2005). Lines that are too close together have been consistently observed to perform less efficiently. At line spacings higher than 2 is where this correlation is lost. Another factor in which there's agreement is font size. Increasing the size is good up until a threshold. There's debate on where such threshold is, with some claiming it sits around 22 points (Rello et al., 2016), and others at 14 (Bhatia et al., 2014).

Measuring readability

Long has passed since typographers started intuitively optimising paragraph presentation for more efficient reading. As mentioned, recent academic research - most of which has happened within the last 50 years, has built a corpus which strongly defines how the readability of a text can be measured.

We consider readability as the quality of a text for being easily understandable by a reader. Readability depends on both the content of the text and its presentation. For the content, several metrics have been used to measure the ease of reading of a certain text. Two highly replicated ones are by Flesch (1943 and 1948), which consider syntactic values (length of sentences, number of words and average syllables per word), to provide with a score. For the purposes of this study, the traditional Flesch Reading Ease Score and the 1970s reformulated Flesch-Kincaid grade-level score were used as measuring tools for text complexity.

With the complexity of the content controlled by these scores and presentation being the experimental levels studied, other variables can be used to accurately measure ease of reading or readability. One of these insightful variables is reading speed. Academia has shown texts with the same length will be read quicker if they're easier to read (Rayner et al., 2016). Very high speed of reading, though, is linked with a decrease in accuracy (Rayner et al., 2016), which translates to reduced comprehension. For such, measuring accuracy in reading is also necessary. A way of doing so is by checking for comprehension by asking questions about the text.

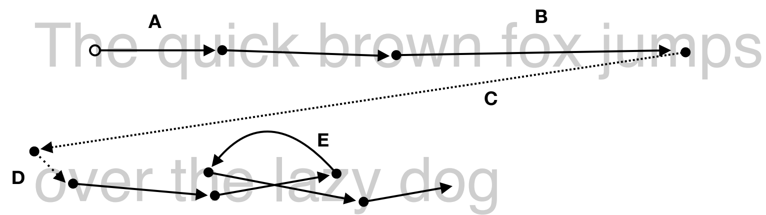


Figure 8. Representation of the expected eye tracker data showing eye movements of interest to reading research.

A. Forward saccade; B. Skip; C. Return sweep; D. Refixation; E. Regression.

Based on Rayner et al. (2016)

Eye movement readings are another very insightful metric to understand readability (see Figure 8 for relevant reading-related eye movements). When reading paragraph text, reading patterns show cognitive insight about how well the reader is understanding the text. A text that's more readable usually requires less regression saccades. These are re-reading saccades, returns to the previous word or few words. The amount of these is inversely proportional to the readability of the text (Rayner et al., 2016).

Another eye tracking metric that correlates with readability are return sweeps. Return sweep is the name given to the eye movement that happens when advancing to the next line of text (see C in Figure 6). More accurate back sweeps are linked with comfort in reading (Parker et al., 2019). These eye movements are usually guided by parafoveal processing, for which the saliency and differentiation between lines are factors that directly affect return sweeps (Vonk et al., 2000). Fixations are the moments in which the eyes are stationary between saccades. When it comes to these, a text that requires more fixations is usually one that's less readable. Words are usually skipped at a higher rate with simpler texts (Rayner et al., 2016). How long fixations last is also insightful, as simpler texts have smaller average fixation durations (Rello et al., 2013; Atvars, 2017).

In conclusion, to answer the questions of "what variables should be studied", based on how academia has approached readability, the following variables were measured: speed of reading, comprehension, amount of returns to re-read, fixation count and length of fixations.

Rationale

The agreement there is about the impact spacing and font size have on readability can't be observed with the aforementioned repetition of words. Little research has been conducted on tools that help keeping track of line position, apart from some commercial endeavours like BeeLine Reader and Bionic Reading. Given that word repetition has been historically used to ease navigation and literature shows it does not negatively affect readability, this resource seems like a potentially interesting tool to help readers follow a reading pattern.

Additionally and following more recent reading research, which suggests line differentiation helps readability, another addition was studied. This addition consisted of diamond shapes to cue the reading of each line. Unlike word additions, which require a higher level of processing, these shapes are simple enough to be clearly perceived in the periphery (Sandberg et al., 2008), guiding eye movements between lines.

Research questions

With the previous background and rationale, a series of research questions were posed for the purpose of the present research. A big research question was then broken down in to other smaller ones, which fall into two categories:

- How do additions before each line affect readability?
    - ***How do word repetitions affect readability?***
        - Is there a difference between outside and embedded words?
    - ***How do graphic additions (diamond shapes) affect readability?***

Methods and procedure

Pilot

A pilot study was conducted to find the ideal parameters for the experiment. Using the same EyeLink setup, three parameters were studied: width of lines, line spacing and font size. Stimuli consisted of the same passage from the ugly duckling in different levels of the three parameters. The levels were determined using previous literature (see Figure 9 for parameter levels). What these papers considered to be the ideal width, line spacing and font size was used to generate the following levels. These were translated to the values used by the 1080x1920 pixel screen present in the EyeLink setup and used for the entirety of this research. The levels studied during the pilot were:

Width
- Small width (676 pixels)
- Large width (948 pixels)

Line spacing
- Low spacing (7.5 points)
- Medium spacing (10 points)
- Large spacing (12.5 points)

Font size
- Small font size (4.5 points)
- Large font size (5.5 points)

Figure 9. Parameters used for the pilot study. Based Rello et al., 2016 - *The Effect of Font Size and Line Spacing on Online Readability.* and Nanavati and Bias, 2005 - *Optimal Line Length in Reading--A Literature Review.*

Twelve blocks of text with all possible style combinations were produced in Helvetica Bold. On the left margin of each block and in the same style as the rest of the text, the last word of the previous line was repeated at the beginning of the next. Text was justified to the right to reduce line differentiation. All stimuli was presented on a black background in white with 30% opacity for body text and 15% opacity for the additions to reduce afterimage.

Three participants were asked to read the twelve passages, which were shown to each in a different order. Participants pressed the space bar to move on to the next text, which allowed to measure reading time. Using EyeLink fixations and left saccades were counted, which have been shown to indicate ease of reading (Slattery & Parker, 2019). Additionally, participants were asked to verbally rate ease of reading from 1 to 10 after each text.

Results from the pilot study suggested the ideal parameters for this setup are a line spacing of 10 points, 5.5 points of font size and a line width of 948 pixels, which resulted in around 35 characters per line. This decision was taken from the qualitative data collected. Two of the three participants rated the aforementioned parameters the highest out of the twelve alternatives, with a score of 9 and 10 respectively. Comparing reading time, fixations and left saccades didn't return any conclusive result. Three-way and two-way repeated measures ANOVAs were performed and no significant differences were found between the twelve levels. Only larger font size seemed to perform better than smaller ($p=0.12$). Having set the ideal text style parameters, the experiment was conducted using similar methods.

Study methods

An eye-tracking experiment was built to measure the parameters aimed to answer the research question of this study. The English translation of the Little Prince (Antoine de Saint-Exupéry, originally published in 1943) by Richard Mathews was used for the stimuli (*The Little Prince*, 2020). A set of 32 passages was extracted from the book. A comprehensibility without context was given to each. By using this parameter together with text length (smallest distance to 275 characters), the passages were ranked. The twenty first scoring ones were selected. No text was longer than 300 characters or shorter than 250.

Additionally, all passages were measured using the Flesch-Kincaid Grade Level and Flesch Reading Ease Score (Flesch, 1943; Flesch, 1948), indicators of ease of reading. All twenty scored between grade 5 and 8 in the Flesh-Kincaid test (middle-school level) and above 75 in the Flesch test. Three texts were altered to fit the criteria mentioned. Two were added an adjective to reach the character count of 250 and one had its subject changed. See Appendix A for the texts used in the experiment.

Twenty multiple-choice questions were created to measure comprehension. They referred to the subject, tone or overall message of the text or to a specific detail mentioned in the passage. Each question was joined by four A-D choice answers. The full questionnaire is under Appendix B.

For the text additions, two versions of these were selected: Dimmed and separated from the body text and embedded in the block. Another condition was added to measure the effect of graphic additions. A diamond shape in the same colour of the text was added before each line, in the hopes it would help parafoveal processing. Additionally, a control condition with no alterations before the block of text was used. See Figure 10 for reference.

Text addition conditions

No text conditions

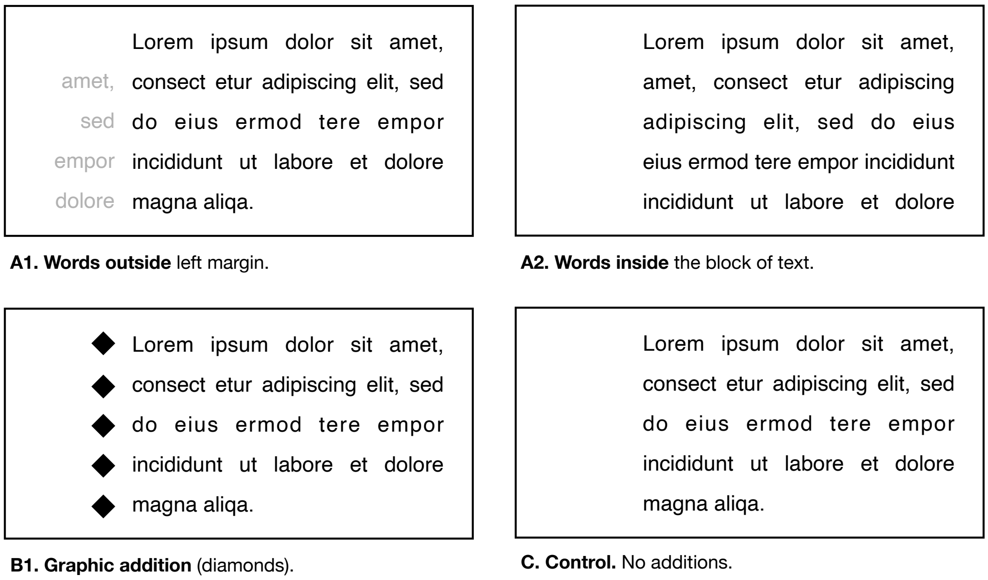


Figure 10. The four conditions used in the experimental design of this research using placeholder text for presentation purposes. Participants read meaningful pieces of text.

The twenty texts were presented to all participants in a random order, following an A, B, C, D pattern for the conditions. This was to avoid training effect. Before the presentation of each stimuli, participants saw a fixation symbol where the first letter of the paragraph would later appear. Five instruction slides were designed and participants were walked through them before starting. Each participant also took a trial round before the beginning of the experiment. See Appendix C for the full set of slides.

All stimuli was generated using Adobe Illustrator and Photoshop 2022. Images were exported in PNG format with a resolution of 1080x1920 pixels. OpenSesame version 3.3.11 was used to put the experiment together. PyGaze components were added to the workflow to record and store eye-tracking data during stimuli presentation. The eye-tracker used was an SR Research EyeLink 1000, which was arranged below the stimuli screen. Participants were sat in a movable chair and placed their chin and forehead on a rest at around 75 centimetres from the eye-tracker's camera. To improve the quality of the readings, the walls of the experiment room were black and the overhead light was dimmed. On the table, a Windows 10 desktop computer executed the experiment while the experimenter sat behind the participant and recorded the comprehension responses.

Before the experiment started, qualitative data was collected about the relationship each participant had had with reading. The following questions were used:

- How was your experience of learning how to read?
- How good of a reader were you during primary school?
- Did you like reading when you were younger? Do you still read for fun?
- Nowadays, how often do you read fiction novels? What about academic texts?
- What's your experience with reading out loud?

Participants

Twenty participants were recruited mainly from the cohorts of MSc Applied Cognitive Psychology and BSc Psychology at Utrecht University. Eight participants came from different backgrounds. Convenience recruiting as well as SONA systems were used to reach participants.

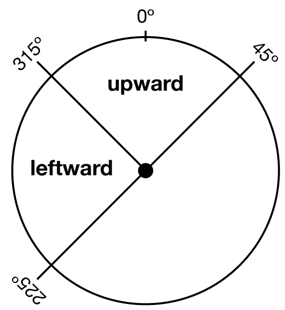
Participants ranged 19 to 29 in age (avg. 23.5), and the sample was comprised of 13 women, 6 men and one non-binary person. All participants were proficient in English, and had corrected or perfect vision. Three of these declared having ADHD, and one was diagnosed with hyperactivity. No other diagnosed learning disabilities were reported.

Results

The data obtained from the experiment consisted of 20 eye-tracker recordings of 20 trials each, the response each participant gave to all questions about the passages, qualitative data about the participant's relationship with reading and some other basic demographic information about them. The eye-tracker's output was generated at a frequency of a thousand recordings per second in EDF file format. To interpret this data, SR Research's software EyeLink Data Viewer was used. The version of the software was 1.6.4.

The variables interpreted by the eye-tracker data software were:

- Duration of the longest fixation from each trial.
- Amount of leftward saccades per trial (regression saccades).
- Amount of upward saccades per trial (regression saccades).
- Amount of fixations per trial.



Direction of saccades was determined based on the angle their vector produced (see left). Saccades between 315 and 45 degrees were considered upward, and those between 225 and 315 leftward.

The data from these variables was organised and preprocessed using Microsoft Excel 16.49. The participants' answer to each trial's question was added to the aforementioned variables measured. JASP version 0.14.1 was used to analyse the data and produce the following results.

Speed of reading

Participants were asked to press the space bar when they were done reading each of the texts. The time from stimuli onset until space bar pressing was measured. The time each participant took to read all five texts from each condition was averaged, and these were compared between conditions. See Figure 11 for these results.

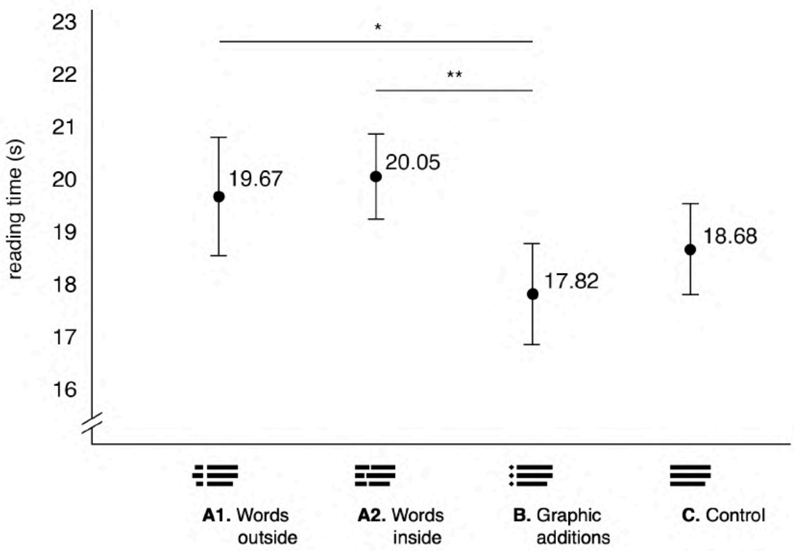


Figure 11. Average reading time per trial per condition in seconds. Horizontal lines represent significant difference between averages. p≤0.01 is represented with two asterisks, p≤0.05 with one.

Another four-way repeated-measures ANOVA was performed to measure the difference between the four conditions. The average reading time for conditions B and C was lower than A1 and A2, but the difference between A2 and B ($p<0.01$) and A1 and B ($p=0.02$) were the only two statistically significant.

Amount of returns to re-read

Upward and leftward saccades were counted to measure how many times participants returned to re-read in each trial (see Figure 12).

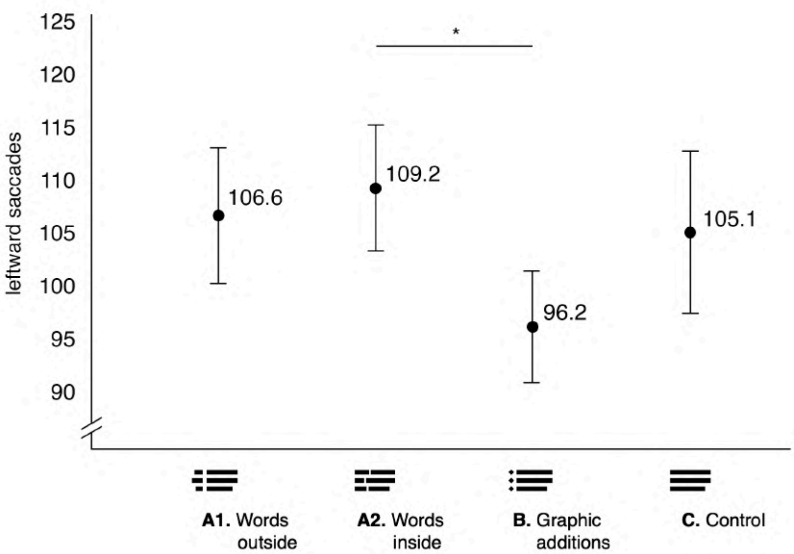


Figure 12. Average count of leftward saccades per trial per condition. Horizontal lines represent significant differences. One asterisk represents p≤0.05.

The five trials each participant completed in each of the four conditions were averaged and a four-way repeated-measures ANOVA was used to compare the amount of leftward saccades between conditions. A significant difference was found between conditions A2 and B (p=0.02).

A similar procedure was carried out for upward saccades. ANOVA studies showed no significant difference could be found between conditions (see Appendix D for graph with these results).

Fixations

Two more variables were considered to measure ease of reading: Fixation count per trial and duration of the longest fixation (see Figure 13).

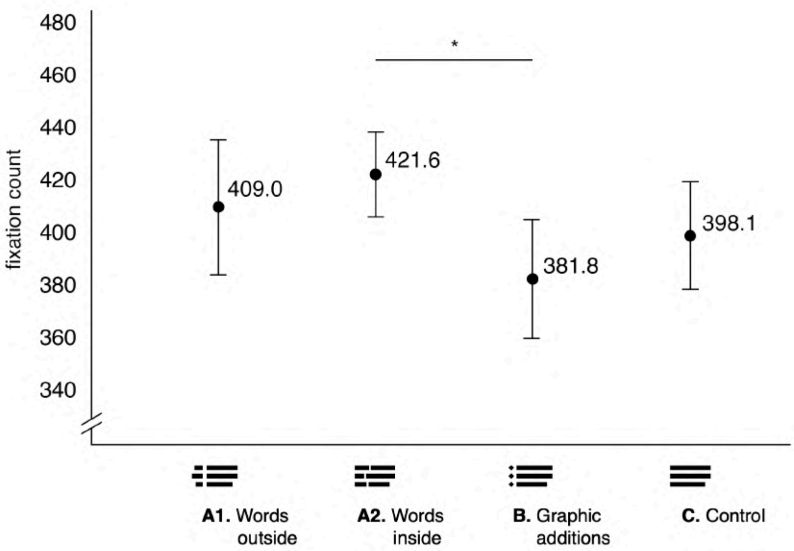


Figure 13. Average count of fixations per trial per condition. Horizontal lines represent significant differences between averages. One asterisk represents $p \leq 0.05$.

Results showed *B. Graphic additions* scored the lowest in fixation count, with an average of 381.8 fixations per trial, while *A1. Words outside* was the one with the largest amount of fixations, scoring at 421.6. Another within-participants ANOVA study was performed and post hoc results showed a significant difference between conditions A2 and and B.

Additionally, longest fixations were plotted and compared, and *A1. Words outside* showed the highest scoring of these, with 2.87 seconds. The shortest longest fixation average was produced by *B. Graphic additions*. ANOVA results showed no significant difference between conditions. See Appendix E for results.

Comprehension

Responses to the multiple choice questions were recorded. Every participant was shown five questions from each condition, for which data consists in a whole number from 0 to 5, being 0 no correct answers and 5 all questions answered correctly.

A four-way repeated-measures ANOVA was performed to measure the difference between conditions. Although the average for the two control conditions was higher (4.25 for *B. Graphic additions* and 4.20 for control), the post hoc test returned no significant difference between any two conditions. See Appendix F for a graph on these results.

Other results

The data was divided in participants with and without ADHD or hyperactivity, and analysis of variance was performed between both in the six variables studied. ANOVA studies returned no significant difference between ADHD and neurotypical participants for any of the conditions.

Qualitative responses were also used to compare the data. Participants that reported having recent experience reading fiction and were good at learning how to read were compared against those who reported the opposite. No conclusive results were found in any of the six variables studied.

Discussion

Results suggest out of the four alterations studied, *B. Graphic additions* was the most helpful when reading. Data shows reading time, leftward saccades and fixation counts were significantly lower than those of other conditions. Although results from these conditions weren't significantly different against control, the trend observed suggests an improvement when using *B. Graphic additions* against other additions.

The poor results of *A1. Words outside* suggest that, at least without training, anadiplosis is not ignored by readers. The effect is even worsened when the words are embedded in the text, as results of *A2. Words inside* show. This aligns with prior knowledge of return sweeps being guided by parafoveal information, rather than foveal (Slattery & Parker, 2016). When readers are nearing the end of a line, their peripheral perception of the beginning of the next line is not defined enough to provide useful information, resulting in it being useless for navigation.

Additionally, when readers completed the return sweep and landed on the beginning of the new line, they encountered the repeated word and had to spend cognitive effort in processing it again and perhaps ignoring it. Looking at the eye-tracking readings, several participants started ignoring the repeated words of *A1. Words outside.* This did not happen in the *A2. Words inside* condition, which didn't allow for easy discrimination of the repeated words.

The good results of *B. Graphic additions* (see Figure 14), though, are certainly remarkable. Participants were faster, spent less fixations per trial and re-read less when diamonds were present at the beginning of each line. This was probably caused by the increased

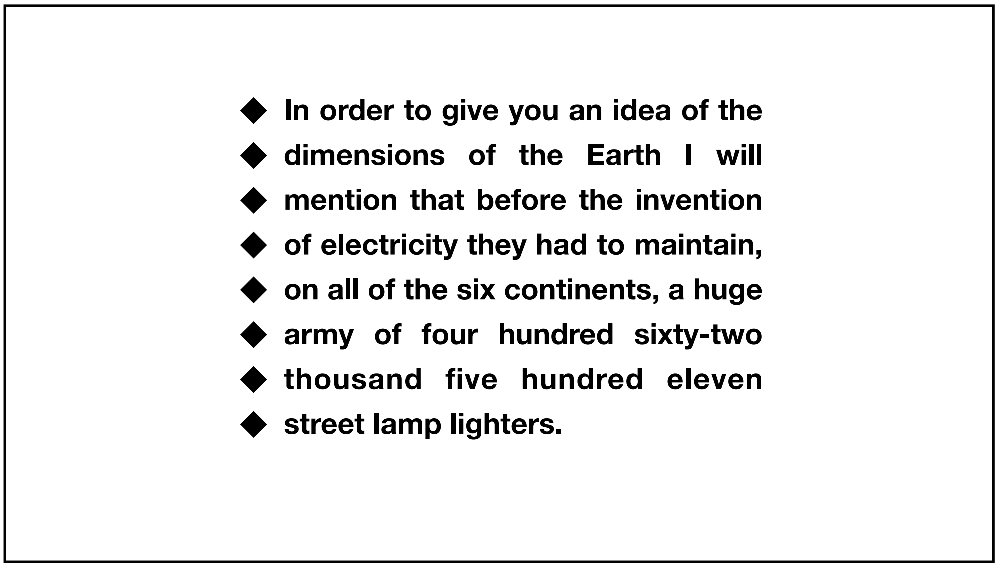


Figure 14. Presentation of text 4 from the experimental stimuli using condition *B. Graphic additions* (see Appendix A for all texts used). Colours inverted for presentation purposes.

differentiation between lines these diamonds offer. Unlike conditions *A1* and *A2*, where words weren't useful, as they weren't helpful at all before the return sweep had occurred, the diamonds in this condition remained visible in the periphery throughout. Participants had enough acuity in the parafoveal area to discern the diamonds when finishing each line. This allowed quicker transitions between lines, easing reading.

These findings align with the rest of literature, which suggests return sweeps are guided by larger perceptual elements of the layout of the text, such as the shape well differentiated lines project within the paragraph (Bhatia et al., 2014; Rello et al., 2016. See Figure 15). Larger separation between lines has been observed very helpful to improve reading ability, as lines are more easily clustered. Larger initials, which also help discern between lines, are also helpful (Rello et al., 2016). For the first time though, the preset findings show additions unrelated to the text's design, also help.

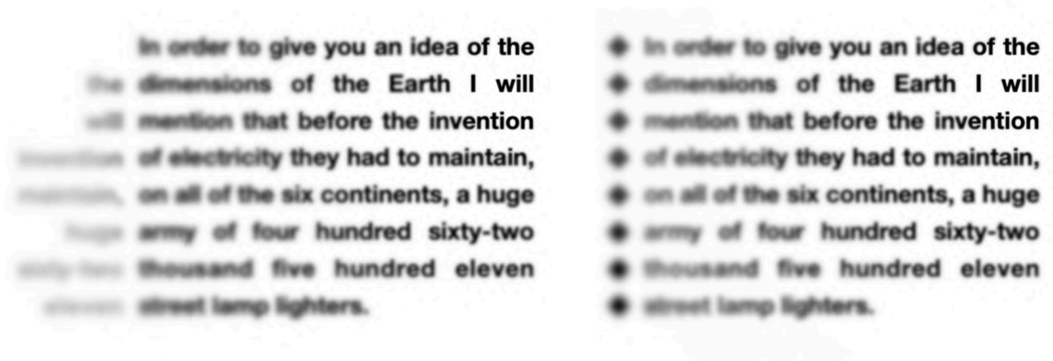


Figure 15. Representation of the visual acuity of a reader when focusing on the word "will", at the end of the second line. Based on Tan, Weimin & Yan, Bo. (2017). The further from the foveal point, the less acuity. Comparing this representation using *A1. Words outside* and *B. Graphic additions* respectively shows the advantage of the diamonds, which remain more salient.

Further research should look at the impact of different additions, and how these could be integrated in layouts usable by designers. Different shapes might impact readability in different ways, and the effect of factors like distance, size and overall position should be studied.

The results, although useful to guide further research, can't be used to extrapolate solid generalisable conclusions. Follow up studies should look at the effects of training. Given the design of this research, participants weren't able to anticipate the additions, and were provided with no time to get accustomed to each variation. It can be expected that a reader that's used to a certain addition will find themselves more comfortable reading such text after a while.

The lack of control over the familiarity of the words for conditions A1 and A2 could have been another limitation of the present design. The texts were standardised and controlled for complexity, and the experiment's design - within subjects - reduced potential issues with individual differences. Perhaps a more controlled set of stimuli, with the words in the leftmost additions being high or low frequency, could return different results (aligning with previous research on high-frequency words being skipped more frequently than low frequency words when matched on length, Rayner et al., 1996). Additionally, overall contrast of the additions should be controlled. The present research used faded colours for condition *A1. Words outside* (reduced opacity, 15% against 30% of the text), while the diamonds of *B. Graphic additions* had the same opacity as the text, and occupied more pixels in the screen, resulting in more contrasted blobs.

Nevertheless, the findings of this research provide new insight on how elements surrounding the text affect readability. It also adds a new perspective on readability research, as most literature has looked at modifications within the text (such as typeface, font size, colour of the letters, contrast between letters and line contrast). A new line of readability research could be developed after this study, which has provided new insights that can be usefully applied to future design practice.

References

Atvars, A. (2017). Eye movement analyses for obtaining readability formula for Latvian texts for Primary School. Procedia Computer Science, 104, 477–484. https://doi.org/10.1016/j.procs.2017.01.162

Bhatia, S. K., Samal, A., Rajan, N., & Kiviniemi, M. T. (2011). Effect of font size, italics, and colour count on web usability. International Journal of Computational Vision and Robotics, 2(2), 156. https://doi.org/10.1504/ijcvr.2011.042271

Bisson, M.-J., Van Heuven, W. J., Conklin, K., & Tunney, R. J. (2012). Processing of native and foreign language subtitles in films: An eye tracking study. Applied Psycholinguistics, 35(2), 399–418. https://doi.org/10.1017/s0142716412000434

Brozo, W. G., Schmelzer, R. V., & Spires, H. A. (1983). The Beneficial Effect of Chunking on Good Readers' Comprehension of Expository Prose. Journal of Reading, 26(5), 442–445. http://www.jstor.org/stable/40034503

Castello, E. (2009). Text complexity and reading comprehension tests. https://doi.org/10.3726/978-3-0351-0276-5

Chun, K.-wai, C. (n.d.). Effects of text structure-based knowledge and strategies on second language expository prose comprehension. https://doi.org/10.5353/th_b3124159

Drieghe, D., Slattery, T. J., Liversedge, S. P., & Rayner, K. (2010). Eye movements and word skipping during reading: Effects of word length and predictability. PsycEXTRA Dataset. https://doi.org/10.1037/e520592012-012

Drieghe, D., Slattery, T. J., Liversedge, S. P., & Rayner, K. (2010). Eye movements and word skipping during reading: Effects of word length and predictability. PsycEXTRA Dataset. https://doi.org/10.1037/e520592012-012

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

Educación - mineduc.gob.gt. (n.d.). Retrieved July 5, 2022, from https://www.mineduc.gob.gt/DIGECADE/documents/Telesecundaria/Recursos%20Digitales/3o%20Recursos%20Digitales%20TS%20BY-SA%203.0/PROYECTOS%20INTEGRADOS/U12%20proyecto%2012%20educaci%C3%B3n.pdf

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. Journal of Verbal Learning and Verbal Behavior, 20(6), 641–655. https://doi.org/10.1016/s0022-5371(81)90220-6

Enos, R. L. (2006). The emergence of a literate rhetoric in Greece. Rhetoric Society Quarterly, 36(3), 223–241. https://doi.org/10.1080/02773940500509144

Evaldo Leal, S., Munguba Vieira, J. M., dos Santos Rodrigues, E., Nogueira Teixeira, E., & Aluísio, S. (2020). Using eye-tracking data to predict the readability of Brazilian Portuguese sentences in single-task, multi-task and

Sequential Transfer Learning Approaches. Proceedings of the 28th International Conference on Computational Linguistics. https://doi.org/10.18653/v1/2020.coling-main.512
EyeLink 1000 plus technical specifications - SR research. (n.d.). Retrieved July 5, 2022, from https://www.sr-research.com/wp-content/uploads/2017/11/eyelink-1000-plus-specifications.pdf

Flesch, R. (1943). Marks of Readable style. Teachers College Record: The Voice of Scholarship in Education, 45(6), 1–2. https://doi.org/10.1177/016146814404500606

Flesch, R. (1948). A new readability yardstick. Journal of Applied Psychology, 32(3), 221–233. https://doi.org/10.1037/h0057532

Hamel, C. D. (1995). Scribes and illuminators. Univ. of Toronto Press u.a.

Hooge, I. T., Holleman, G. A., Haukes, N. C., & Hessels, R. S. (2018). Gaze tracking accuracy in humans: One eye is sometimes better than two. Behavior Research Methods, 51(6), 2712–2721. https://doi.org/10.3758/s13428-018-1135-3

Hooge, I., & Camps, G. (2013). Scan path entropy and arrow plots: Capturing scanning behavior of multiple observers. Frontiers in Psychology, 4. https://doi.org/10.3389/fpsyg.2013.00996

Houston, K. (2013, September 25). The rise and fall of the Pilcrow, part I. Slate Magazine. Retrieved July 5, 2022, from https://slate.com/human-interest/2013/09/the-pilcrow-how-the-paragraph-punctuation-mark-evolved-from-ancient-greece-and-rome.html

Kelih, E., & Mačutek, J. (2011). Harald Baayen. 2008. analyzing linguistic data. A practical introduction to statistics using R. Cambridge: Cambridge University Press. XII, 353 s. Zeitschrift Für Rezensionen Zur Germanistischen Sprachwissenschaft, 3(2). https://doi.org/10.1515/zrs.2011.035

Kennedy, A. (2016). Book review: Eye tracking: A comprehensive guide to methods and measures. Quarterly Journal of Experimental Psychology, 69(3), 607–609. https://doi.org/10.1080/17470218.2015.1098709

A KÓDEX: Magyar Nyelv: Sulinet Tudásbázis. Tudásbázis. (n.d.). Retrieved July 5, 2022, from https://tudasbazis.sulinet.hu/hu/magyar-nyelv-es-irodalom/magyar-nyelv/magyar-nyelv/nyelvemlekeink/a-kodex

Lai, M.-L., Tsai, M.-J., Yang, F.-Y., Hsu, C.-Y., Liu, T.-C., Lee, S. W.-Y., Lee, M.-H., Chiou, G.-L., Liang, J.-C., & Tsai, C.-C. (2013). A review of using eye-tracking technology in exploring learning from 2000 to 2012. Educational Research Review, 10, 90–115. https://doi.org/10.1016/j.edurev.2013.10.001

Leckner, S. (2012). Presentation factors affecting reading behaviour in readers of Newspaper Media: An eye-tracking perspective. Visual Communication, 11(2), 163–184. https://doi.org/10.1177/1470357211434029

Lewis, E. H. (1974). The history of the English paragraph. Library of Congress Photoduplication Service.

Low, D. V. T. (1901). The practice of typography: Correct composition: A treatise on spelling, abbreviations, the compounding and division of words, the proper use of figures and numerals ... with observations on punctuation and proof-reading. Century Co.

Lunsford Guide Instructor Notes - DePaul University. (n.d.). Retrieved July 5, 2022, from https://condor.depaul.edu/dwrd/files/lunsford-guide-instructor-notes.pdf

Nanavati, Anuj & Bias, Randolph. (2005). Optimal Line Length in Reading--A Literature Review. Visible Language.

Newspaper Column Inches - Archive.org. Newspaper column inches. (n.d.). Retrieved July 5, 2022, from https://web.archive.org/web/20121217014138/http://graphicdesign.spokanefalls.edu/tutorials/tech/column_inches/col_inches.htm

Niehorster, D. C., Cornelissen, T. H., Holmqvist, K., Hooge, I. T., & Hessels, R. S. (2017). What to expect from your remote eye-tracker when participants are unrestrained. Behavior Research Methods, 50(1), 213–227. https://doi.org/10.3758/s13428-017-0863-0

Parker, A. J., Slattery, T. J., & Kirkby, J. A. (2019). Return-sweep saccades during reading in adults and children. Vision Research, 155, 35–43. https://doi.org/10.1016/j.visres.2018.12.007

Parkes, . (1993). A social prehistory of European languages. Antiquity, 67(257), 747–760. https://doi.org/10.1017/s0003598x00063766

Raney, G. E., Campbell, S. J., & Bovee, J. C. (2014). Using eye movements to evaluate the cognitive processes involved in text comprehension. Journal of Visualized Experiments, (83). https://doi.org/10.3791/50780

Rayner, K., Schotter, E. R., Masson, M. E., Potter, M. C., & Treiman, R. (2016). So much to read, so little time. Psychological Science in the Public Interest, 17(1), 4–34. https://doi.org/10.1177/1529100615623267

Rayner, K., Schotter, E. R., Masson, M. E., Potter, M. C., & Treiman, R. (2016). So much to read, so little time. Psychological Science in the Public Interest, 17(1), 4–34. https://doi.org/10.1177/1529100615623267

Readability: The optimal line length – articles. Baymard Institute. (n.d.). Retrieved July 5, 2022, from https://baymard.com/blog/line-length-readability

Readability: The optimal line length – articles. Baymard Institute. (n.d.). Retrieved July 5, 2022, from https://baymard.com/blog/line-length-readability

Rello, L., Baeza-Yates, R., Dempere-Marco, L., & Saggion, H. (2013). Frequent words improve readability and short words improve understandability for people with dyslexia. Human-Computer Interaction – INTERACT 2013, 203–219. https://doi.org/10.1007/978-3-642-40498-6_15

Rello, L., Pielot, M., & Marcos, M.-C. (2016). Make it big! The Effect of Font Size and Line Spacing on Online Readability. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. https://doi.org/10.1145/2858036.2858204

Rubin, G. S. (2013). Measuring reading performance. Vision Research, 90, 43–51. https://doi.org/10.1016/j.visres.2013.02.015

Saint-Exupéry, A. de. (2020). The little prince. Arcturus.

Sandberg, M. A., Brockhurst, R. J., Gaudio, A. R., & Berson, E. L. (2008). Visual acuity is related to parafoveal retinal thickness in patients with retinitis pigmentosa and macular cysts. Investigative Opthalmology & Visual Science, 49(10), 4568. https://doi.org/10.1167/iovs.08-1992

Shaikh, A. D., & Chaparro, B. S. (2005). The effects of line length on reading performance of online news articles. PsycEXTRA Dataset. https://doi.org/10.1037/e577292012-014

Slattery, T. J., & Parker, A. J. (2019). Return sweeps in reading: Processing implications of undersweep-fixations. Psychonomic Bulletin & Review, 26(6), 1948–1957. https://doi.org/10.3758/s13423-019-01636-3

Starch, D. (1915). The measurement of efficiency in reading. Journal of Educational Psychology, 6(1), 1–24. https://doi.org/10.1037/h0073433

Starch, D. (1915). The measurement of efficiency in reading. Journal of Educational Psychology, 6(1), 1–24. https://doi.org/10.1037/h0073433

Supplemental material for word frequency, predictability, and return-sweep saccades: Towards the modeling of eye movements during paragraph reading. (2019). Journal of Experimental Psychology: Human Perception and Performance. https://doi.org/10.1037/xhp0000694.supp

Tan, Weimin & Yan, Bo. (2017). Salient object detection via multiple saliency weights. Multimedia Tools and Applications. 76. 1-17. 10.1007/s11042-017-4725-7.

Veldre, A., Reichle, E. D., Wong, R., & Andrews, S. (2020). The effect of contextual plausibility on word skipping during reading. Cognition, 197, 104184. https://doi.org/10.1016/j.cognition.2020.104184

Vonk, W., Radach, R., & van Rijn, H. (2000). Eye guidance and the saliency of word beginnings in reading text. Reading as a Perceptual Process, 269–299. https://doi.org/10.1016/b978-008043642-5/50014-0

West, S. (1963). Reclamantes in greek papyri. Scriptorium, 17(2), 314–315. https://doi.org/10.3406/scrip.1963.3188

Whitford, V., & Titone, D. (2014). The effects of reading comprehension and launch site on frequency–predictability interactions during paragraph reading. Quarterly Journal of Experimental Psychology, 67(6), 1151–1165. https://doi.org/10.1080/17470218.2013.848216

Appendices

	Appendix A. Text fragments from The Little Prince (Antoine de Saint-Exupéry, English translation by Richard Mathews, 2020), stimuli for the experiment	**Appendix B.** Questions to test for comprehension.
1	"It's necessary to demand from each person what they're able to give," answered the king. "Authority rests first on reason. If you order your people to go throw themselves into the sea, there would be a revolution! I have the right to demand obedience because my orders are reasonable."	What character was speaking? A. A king B. A prince C. A rose D. A woman
2	You should never listen to flowers. What you have to do is to look at them and smell them. Mine filled the whole planet with fragrance, but I never knew how to enjoy it. I should have judged her by her actions and not by her words. I should never have fled!	According to the passage, what should you do with flowers? A. Listen to them B. Smell them C. Cut them D. Scream at them
3	"My life is monotonous. I chase the chickens; the men chase me. All the chickens look the same, and all the men look the same. I get a little tired. But if you tame me, my life will brighten up. I will know the sound of footsteps that will be different from all others."	What does the speaker from the text chase? A. Dreams B. Turkeys C. Chicken D. None of the above
4	In order to give you an idea of the dimensions of the Earth I will mention that before the invention of electricity they had to maintain, on all of the six continents, a huge army of four hundred sixty-two thousand five hundred eleven street lamp lighters.	What's the text's goal? A. Give you an idea of amount of workers of a company B. Describe a beautiful place C. Give you an idea of the size of Earth D. None of the above
5	"During the fifty-four years that I have lived on this planet, I have only been disturbed three times. The first time, twenty-four years ago, by a big beetle that fell down God knows from where. It made a frightful noise, and I made four errors in addition."	How many times was he bothered? A. Never B. Once C. Twice D. Three times
6	"You only really know the things you tame," said the fox. "Men no longer have time to really know anything. They buy ready-made things at stores. But since there aren't any stores for friends, men no longer have friends. If you want a friend, tame me! I can be your friend!"	Who was speaking in the previous text? A. A fox B. A dog C. A cat D. A human
7	As the little prince went to sleep, I took him in my arms, and I started walking again. I was moved. It seemed to me I was carrying a fragile treasure. It seemed to me even that there was nothing more fragile on the Earth. In the face of the moon I gazed his pale face, his closed eyes.	How did the prince feel according to the narrator? A. Strong B. Fragile C. Happy D. Sad
8	When it is noon in the United States the sun, everyone knows, is setting in France. You would have to be able to go to France in a minute to be present for the sunset. Unfortunately, France is too far away. But in your tiny planet, it was enough to move your chair a few feet.	What atmospherical event is the text talking about? A. Rain B. Thunderstorm C. Sunset D. Fog
9	The little prince looked around for a place to sit, but the planet was completely covered over by the magnificent cloak the king was wearing. On the planet there was only the king's big throne. The little prince therefore remained standing and, since he was tired, he yawned.	What was covering the planet? A. A coat B. A cloak C. Grass D. None of the above

10	With a sad face, the fox wisely said: “Here is my secret. It's very simple: you only truly see with your heart. What is essential is invisible to the eyes.” "What is essential is invisible to the eyes," repeated the little prince, in order to remember.	According to the text, you can only see with your ____? A. Eyes B. Feelings C. Heart D. Night vision glasses
11	I knew very well that in addition to the great planets -- such as the Earth, Jupiter, Mars, Venus -- to which we have given names, there are also hundreds of others, some of which are so small that one has a hard time seeing them through the telescope.	What topic is the text about? A. Astrology B. Botanic sciences C. Physics D. Art and crafts
12	It’s not the geographer who goes and counts the cities, the rivers, the mountains, the seas and the deserts. The geographer is too important to go wandering around. He never leaves his office. But he welcomes the explorers. He questions them, and he takes notes about their memories.	What profession is described in the text? A. Astronomer B. Biologist C. Geographer D. Cartographer
13	When you tell them about a new friend, they never question you about the basics. They never tell you, “What is the sound of his voice?” What games does he prefer? Does he collect butterflies? They ask, “How old is he? How many brothers? How much does he weigh?"	"People don't ask…” A. The great questions B. The simple questions C. The basics D. The essentials
14	If it is a twig of radish or rose bush, you can let it grow as it wants. But if it is a bad plant, it is necessary to tear the plant immediately, as soon as it is known to recognize it. There were terrible seeds on the planet of the little prince, the baobabs.	What is a bad plant, according to the text? A. The twigs B. The poppies C. The baobabs D. None of the above
15	It’s six years since my friend went away from this planet. If I try to describe it here, it is so I don't forget him. It’s sad to forget a friend. Not everyone has had a friend. And I can become like the grown-ups who are interested only in plain numbers.	How long ago did the friend leave? A. Two weeks B. Two years C. Four years D. Six years
16	In the book it said: “Boa constrictors swallow their prey whole, without chewing. Afterwards they cannot move, and sleep for six months digesting.” I thought a great deal about goings-on in the jungle and, in turn, with a crayon, managed to produce my first drawing.	What animal group did the book describe? A. Snakes B. Insects C. Spiders D. Birds
17	I have serious reason to believe that the planet from which the little prince came is an asteroid known as B-612. This asteroid has only been seen once through the telescope. That was by a Turkish astronomer in 1909. He then shared his discovery at an astronomy congress.	Who spotted the asteroid? A. A wise man B. A Spanish astronomer C. A Turkish astronomer D. The prince
18	“Oh, where I live,” said the little prince, “it’s not very interesting, it’s very small. I have three volcanoes. Two of these volcanoes are active and one is extinct. But one never knows, I have to take care of it.” “One never knows,” repeated the geographer.	How big is the place described? A. Large B. Relatively large C. Relatively small D. Really small
19	“I know a planet where there is a grumpy-faced gentleman. He has never smelled a flower. He has never looked at a star. He has never loved anyone. He has never done anything other than numbers. And all during the day he repeats like you: ‘I am a serious man! I am a serious man!’	What did the man do? A. Smell a flower B. Love someone C. Looked at a star D. None of the above
20	"When you find a diamond that doesn’t belong to anyone, it’s yours. When you find an island that doesn’t belong to anyone, it’s yours. When you have an idea first, you patent it: it’s yours. And I possess the stars since no one before me has ever dreamed of owning them.”	What's NOT mentioned in the text? A. Diamonds B. Ideas C. Spaceships D. Islands

Appendix C. Instruction slides for the experiment.

Figure C1. Slide 1

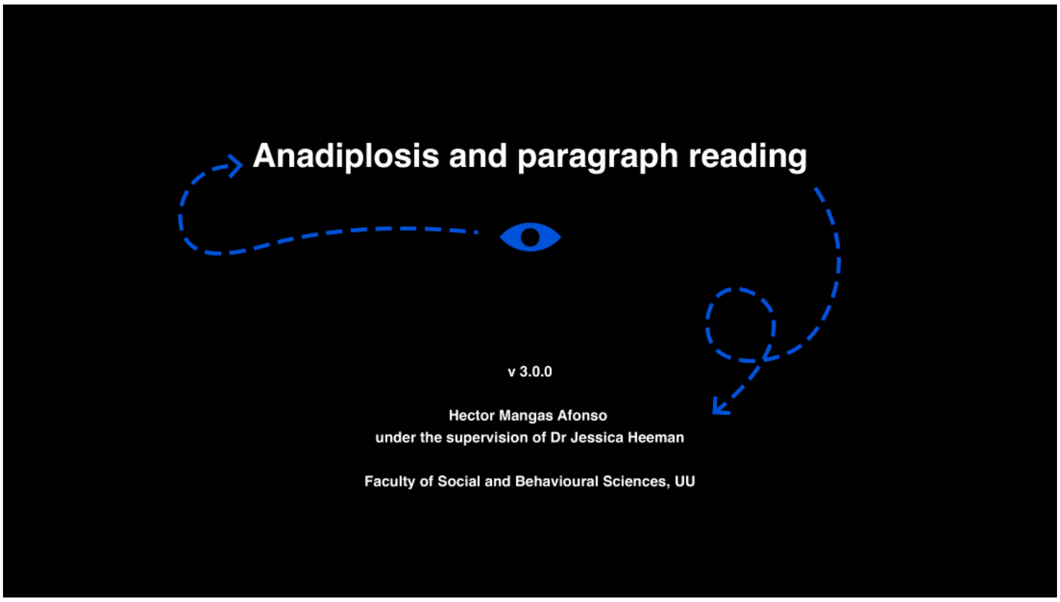


Figure C2. Slide 2

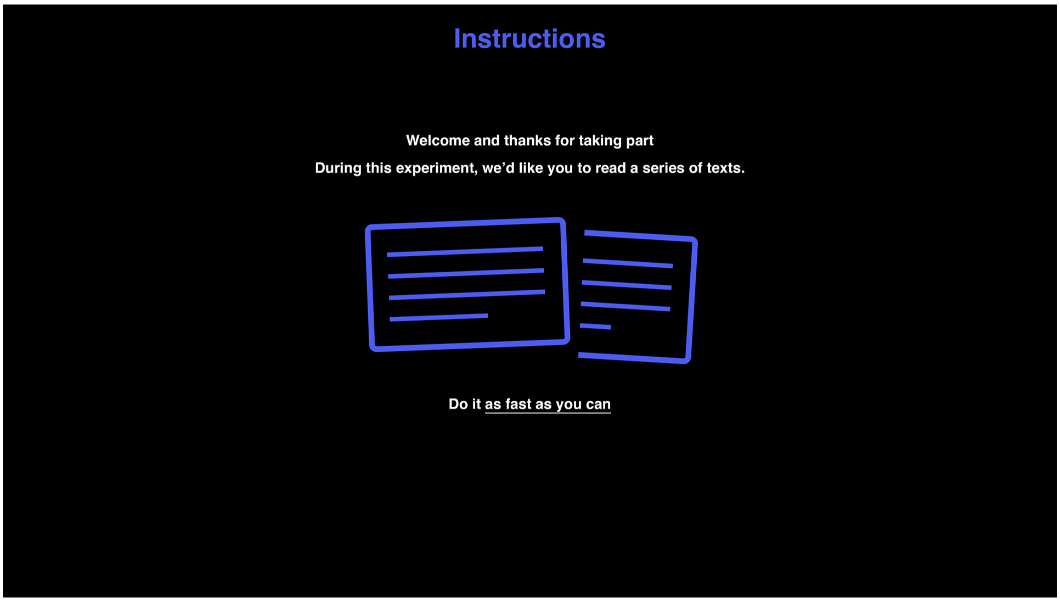


Figure C3. Slide 3

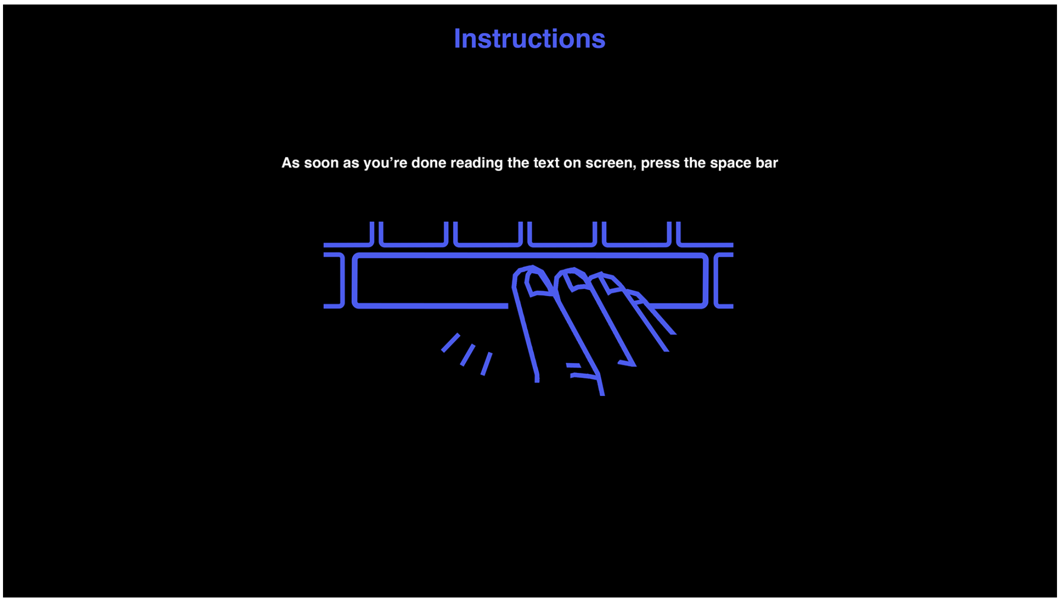


Figure C4. Slide 4

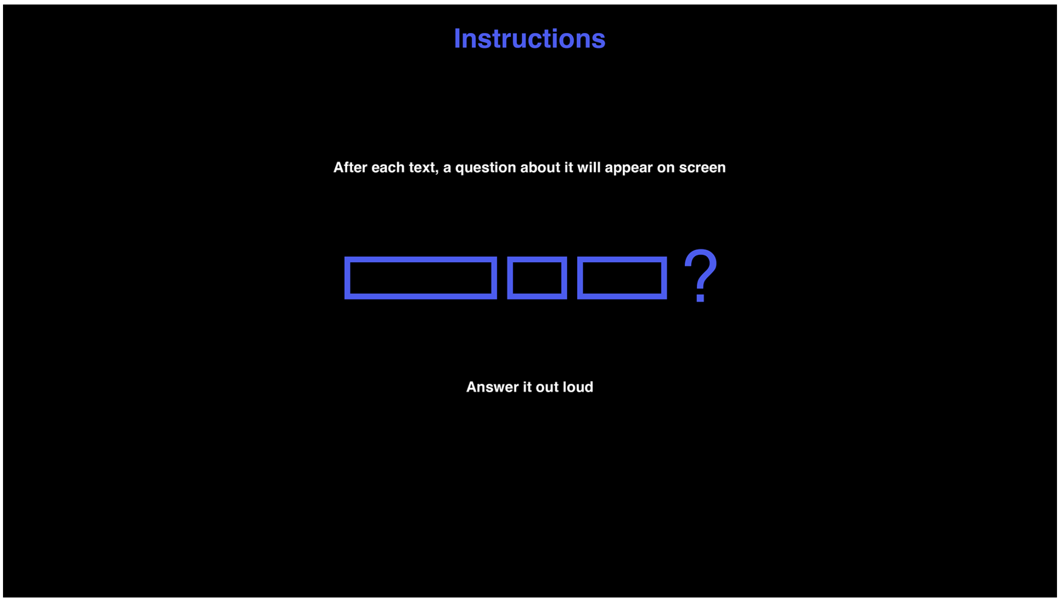


Figure C5. Slide 5

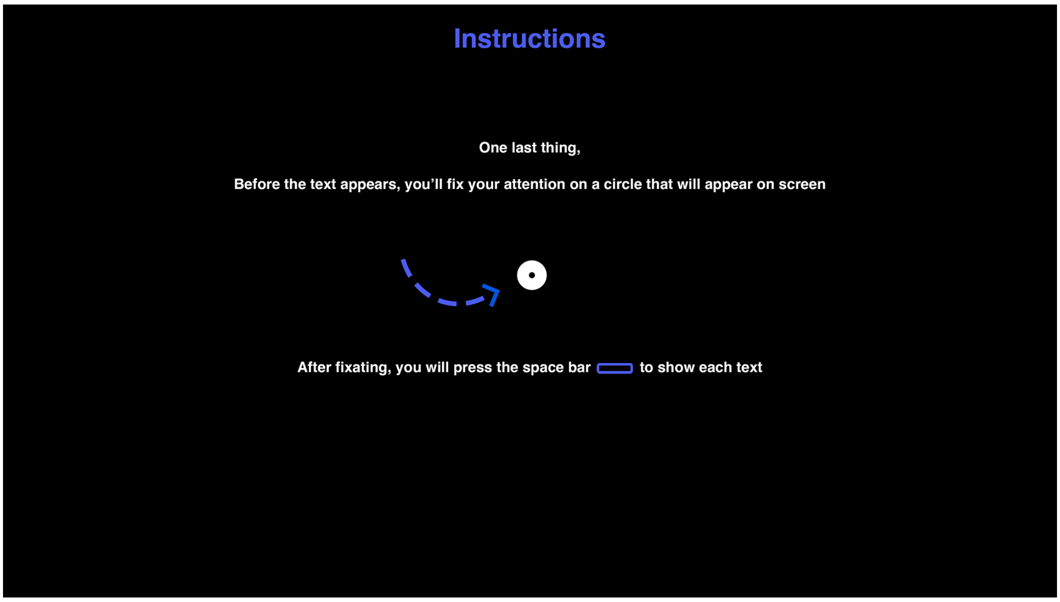


Figure C6. Slide 6

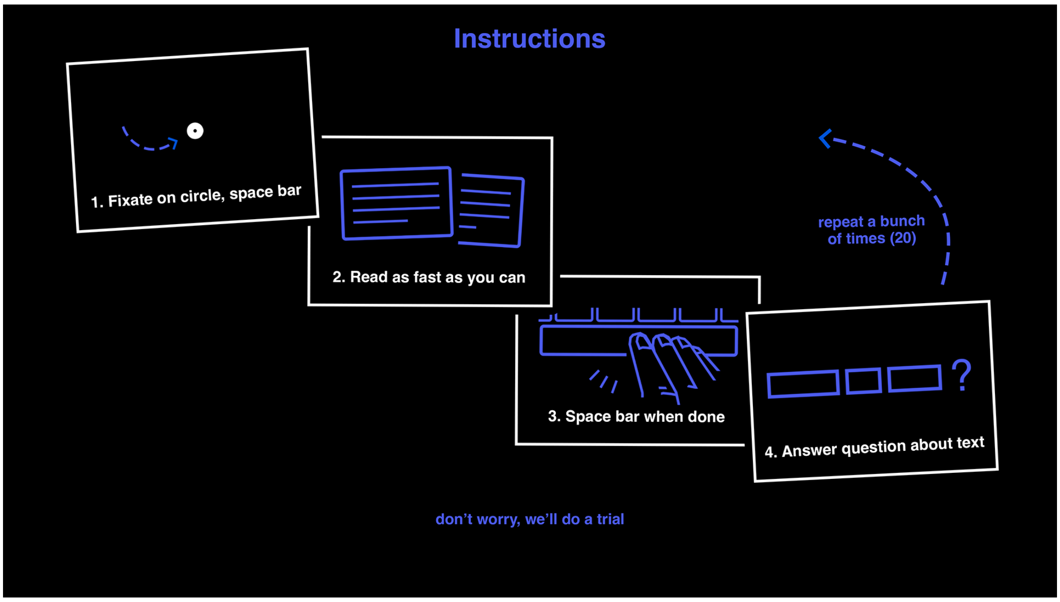


Figure C7. Slide 7

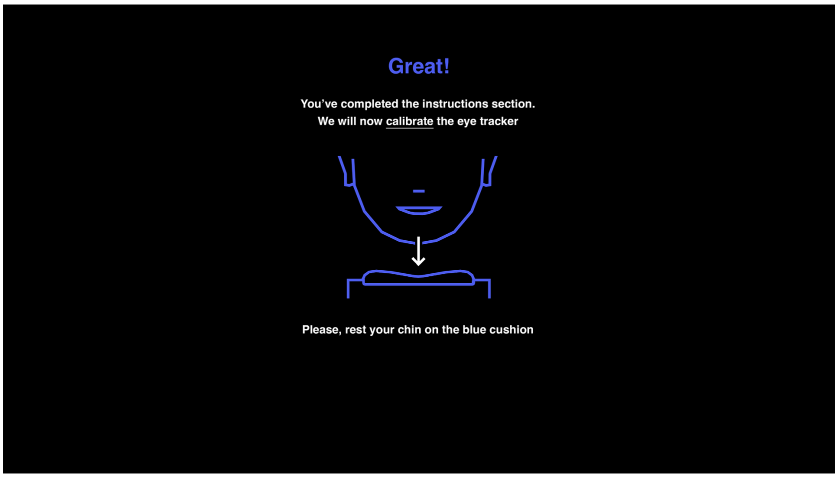
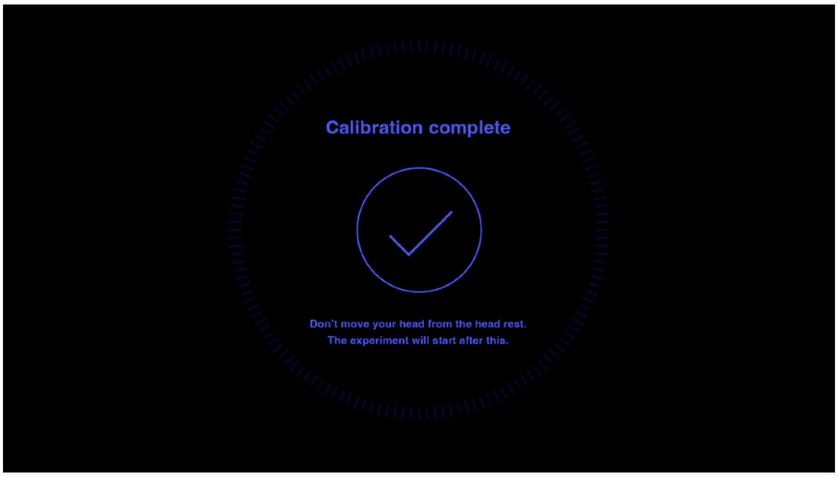


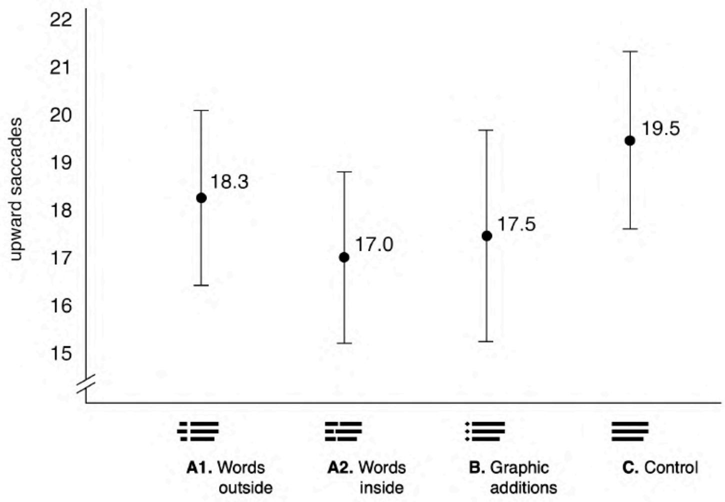
Figure C8. Slide 8

Figure C9. Slide 9

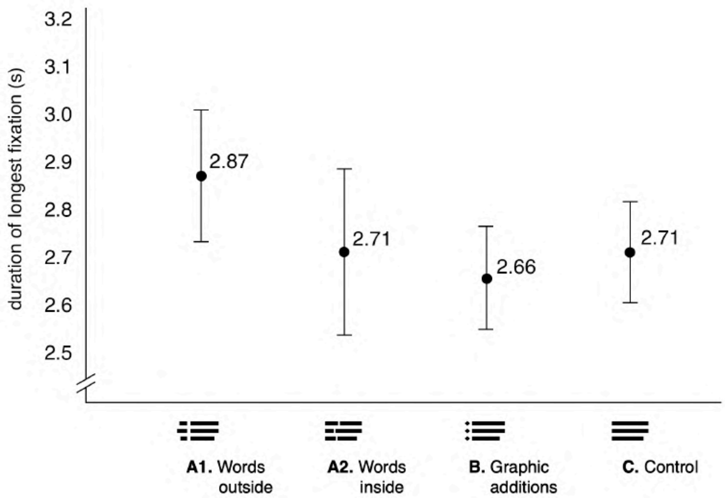
Figure C10. Slide 10



Appendix D. Average count of upward saccades per trial against conditions.



Appendix E. Average duration of the longest fixations of each trial for every condition.



Appendix F. Average correct answers from each participant per condition. Each participant was asked five questions in each condition, for which data ranged 0 to 5 correct answers.