Readability, for whom?

An eye tracking study evaluating the guidelines from Dyslexiförbundet

Kristoffer Karlsson

Supervisor: Arne Jönsson
Examiner: Henrik Danielsson
Copyright

The publishers will keep this document online on the Internet – or its possible replacement – for a period of 25 years starting from the date of publication barring exceptional circumstances.

The online availability of the document implies permanent permission for anyone to read, to download, or to print out single copies for his/her own use and to use it unchanged for non-commercial research and educational purpose. Subsequent transfers of copyright cannot revoke this permission. All other uses of the document are conditional upon the consent of the copyright owner. The publisher has taken technical and administrative measures to assure authenticity, security and accessibility.

According to intellectual property law the author has the right to be mentioned when his/her work is accessed as described above and to be protected against infringement.

For additional information about the Linköping University Electronic Press and its procedures for publication and for assurance of document integrity, please refer to its www home page: http://www.ep.liu.se/.

© Kristoffer Karlsson
Abstract

This work aims to answer the question whether the recommendations for writing a text for a dyslexic reader affects the readability of said text. An eye tracking study was conducted upon texts that were manipulated according to or opposing the guidelines that the Swedish dyslexia association (Dyslexiförbundet) has developed in the project Begriplig text. Participants were 8 dyslexic and 9 non-dyslexic readers. The results show little to no improvement in readability following the guidelines. Future works could examine how the perceived readability corresponds to the results of readability measures, such as the ones calculated by the application TeCST. Future work could also conduct the experiment with more participants.

Keywords: eye tracking, dyslexia, readability, reading comprehension
Acknowledgement

This work has included both eye tracking methodology and statistical analysis of the gathered data. I’d like to thank Linnea Björk Timm for help with the practical issues in the lab, Martin Brengdahl and Ove Jansson for help with the statistics and finally to Janet Eltebo who were my contact at Dyslexiförbundet.

Linköping in June 2018

Kristoffer Karlsson
Table of Contents

1. Introduction ......................................................................................................................... 9
   1.1 About the project ........................................................................................................... 9
   1.2 Purpose .......................................................................................................................... 9
   1.3 Delimitations ............................................................................................................... 9

2. Theory .................................................................................................................................. 11
   2.1 Dyslexia ......................................................................................................................... 11
   2.2 Eye tracking studies ...................................................................................................... 11

3. Method ................................................................................................................................. 13
   3.1 Participants .................................................................................................................... 13
   3.2 Equipment ...................................................................................................................... 13
   3.3 Procedure ....................................................................................................................... 14
   3.4 Design ............................................................................................................................ 15
   3.5 Areas Of Interests .......................................................................................................... 15
   3.6 Measurements ............................................................................................................... 17
   3.7 Readability ..................................................................................................................... 18

4. Results ................................................................................................................................. 19
   4.1 Perceived readability ...................................................................................................... 19
   4.2 Statistics ......................................................................................................................... 19

5. Discussion ............................................................................................................................ 21
   5.1 About the results .......................................................................................................... 21
   5.2 About the method ......................................................................................................... 21
   5.3 Future research ............................................................................................................. 22

6. Conclusion .......................................................................................................................... 23

7. References ........................................................................................................................... 24
List of figures

1. Hyphenation (divided word) ................................................................. 14
2. Area Of Interest covering short sentence ........................................... 16
3. Area Of Interest covering long sentence ............................................. 17

List of tables

1. Participant overview ............................................................................. 13
2. Order of texts ....................................................................................... 15
3. AOI sizes ............................................................................................. 16
4. Number of data collection points .......................................................... 18
5. TeCST scores ....................................................................................... 19
6. Mean values of readability ratings ........................................................ 19
7. Results for the dyslexic group ............................................................... 20
8. Results for the control group ............................................................... 20

List of equations

1. Calculating Cohens d ............................................................................. 19
2. Calculating effect size $r$ for Wilcoxon Signed Rank test ...................... 20
1. Introduction

The following chapter introduces the project from which the guidelines stem and defines the purpose and delimitations for the thesis work.

1.1 About the project

The work is conducted in cooperation with the project “Begriplig text”, and will be performed with the projects coordinator. Begriplig texts purpose is to shed light on what causes the individual’s varying capabilities to comprehend and fully grasp written text, and also to increase the competence and understanding of those responsible for conveying information in written form, or the ones actually writing it. Begriplig text is run in collaboration with the following Swedish associations: the dyslexia association (Dyslexiförbundet), the aphasia association (Afasiförbundet), the Autism and Asperger association (Autism- och Aspergerförbundet), and the association for children and adults with developmental disorder (FUB, Förbundet för unga och vuxna med utvecklingsstörning). Begriplig text is funded by the Allmänna Arvsfonden.

1.2 Purpose

The aim of this study were to evaluate guidelines made by the Swedish dyslexia association, when writing text for a dyslexic reader. This were done by collecting eye tracking data from people with and without dyslexia. Specific areas of interest were a couple of attributes of the texts that had been deemed important or not important by dyslexic participants earlier in the project Begriplig text. These attributes has been collected into a persona made by the dyslexia association.

1.3 Delimitations

This work will not compare the ability to read between individuals and therefore not test the individual’s capacity to read, it will not test how a dyslexic reader fares while reading compared to a non-dyslexic reader. Due to the time limit this work does not aim for a large scale study to be performed.
This work will not take into account age differences nor any possible gender effects.
2. Theory

This chapter will present the theoretical framework that the thesis is built upon.

2.1 Dyslexia

There are several different definitions of dyslexia, and the cause of it is still unknown (Nilsson Benfatto, et al., 2016). However the definition that is going to be used in this paper is the one used by the Swedish dyslexia association, found on their website. It is defined by Ingvar Lundberg (2010) in his book “Läsningsens psykologi och pedagogik”, (the same name as the book by Edmund Huey “The psychology and pedagogy of reading” translated into Swedish) as follows: Dyslexia is a reduction in some linguistic functions. Especially the phonological that are important to use the characteristics of the written languages principles for coding the language. This reduction can first be recognized as difficulties to automatically decode written words while reading, but can also become apparent as difficulties to spell correctly. Secondary consequences can include difficulties with reading comprehension, and limited experience with reading. The reduction is usually appearing in several relatives, and there are cause for believing it has a genetic predisposition. Characteristic of dyslexia is that the reduction is lasting and also intractable. Even though reading with time can reach acceptable levels, the spelling errors usually persists (Dyslexiförbundet, n.d.) There is no relationship between a person’s level of intelligence and dyslexia (Vetenskapsrådet, 2007). Roughly 5-8 % of the Swedish population are believed to have dyslexia (Vetenskapsrådet, 2008)

2.2 Eye tracking studies

Nilsson Benfatto et al. (2016) writes that people in a group with high risk of having dyslexia or actually diagnosed with dyslexia show a different eye movement pattern when reading than people in a group with low risk of having dyslexia do. Fixation lengths are shorter and saccades are longer in the low risk group than it is in the high risk or diagnosed dyslexic group. A fixation is the event where a participant fixates their vision on a stimuli, such as a word, a face or an eye. A saccade is a movement of the participant’s eyes, between their fixations. These movements are often rapid and abrupt (Duchowski, 2007). Eye tracking studies has therefore shown competent in discovering whether or not a person has or is in the risk zone of having dyslexia (Nilsson Benfatto, et al., 2016) Holsanova, Holmberg and Holmqvist found a significant correlation between the time spent reading a text, the order in which they were read, and the number of fixations and the comprehension of the read material when studying the layout of informational graphical elements in newspaper layout (Holsanova, Holmberg & Holmqvist, 2008). According to Holsanova, Rahm and Holmqvist a paratext can be seen as an entry point into the text, they write that Genette describes them as a threshold into the text. Examples of paratexts are headlines, subheadlines, tickers et cetera, but also more global versions, such as reoccurring vignettes or thematic markers. (Holsanova, Rahm & Holmqvist, 2006). During a workshop in Dyslexiförbundets offices in Stockholm, it was noted that the concept of starting of a paragraph with bold style works as a “stepping stone into the text” (Dyslexiförbundet, 2018). Clearly it serves the same purpose as the entry points, or thresholds as Holsanova, Rahm and Holmqvist wrote about (2008).

Using eye tracking equipment to examine reading behavior is nothing new. The articles where Holsanova, Rahm and Holmqvist (2006), and Holsanova, Holmberg and Holmqvist (2008) investigated how a person reads a newspaper spread is one example. Another example is investigating readability. Rello et al. writes that the fixation time can be used as a
measurement of readability, longer fixation duration reading indicate more processing of the data, while shorter fixation durations can mean that the readability is high (Rello et al., 2013).
3. Method

3.1 Participants

A total of 17 participants took part in the study, 8 with dyslexia and 9 without. If the participants did not have adequate eyesight they had the option to correct their eyesight with glasses. The participants were recruited through a mixture between convenience selection and snowball selection, which is participants were asked whether they knew anyone else who could be interested in taking part in the study. The participants were asked beforehand whether or not they had dyslexia, but where not required to provide evidence of the diagnosis in the form of a certificate, however one participant did so anyway.

The result from one participant was not used since the participant used contact lenses leading to erratic looking data. Holmqvist et al writes that soft contact lenses often have a less than perfect fit with the eye and bubbles of air are caught in between the eye and the length. These bubbles cause the infrared light to split up, which causes the eye tracking device to fail to reliably identify the corneal reflection. Instead it reports very rapid eye movements all over the stimuli. These movements are called optic artefacts (Holmkvist et al., 2011).

All participants had secondary school level education or higher. 7 were male and 10 were female (see table 1). All participants gave informed consent verbally after being informed about the experiment, the conditions on which they participated. This was done both verbally and in writing, see procedure.

Table 1. Participant overview

<table>
<thead>
<tr>
<th></th>
<th>Group A – Dyslexic</th>
<th>Group B – Not dyslexic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Female</td>
<td>6</td>
<td>4</td>
</tr>
</tbody>
</table>

3.2 Equipment

The study took place in the eye tracking lab at Linköping University. The room was controlled for infra-red light which can be a source of error (Holmqvist et al, 2011) this was done by inspecting the lights in the ceiling and the windows (that were covered with curtains) to see if they generated, or let any infrared light through. Such was not the case. The equipment used were a SensoMotoric Instrument Remote Eye tracking Device 500, with 500Hz update frequency. The stimuli were showed on a 22 inch screen with the resolution of 1680x1050.

The software used were iView X 500Hz, Experiment Center 3.6 and BeGaze 3.6. All developed by SensoMotoric Instruments.

The stimuli were two texts, taken from the national tests in reading comprehension on secondary school level. They were selected since they were roughly equal in difficulty. One of the texts were manipulated according to guidelines from the dyslexia association, and became the “easy” text. This text was only manipulated visually, with the exception of changing written numbers to letters (“1000” became “one thousand”). The other text was manipulated against the guidelines, the “hard” text. Both texts were cut short at a point where the text could have ended naturally, so neither were longer than a single page.

The experimental manipulations were:

Sentence length, which are deemed as important for the dyslexic reader. The “hard” text had two compound sentences that originally consisted of two separate sentences each. To merge the two sentences into a single compound sentence the texts were examined, and a dot was exchanged for a comma, followed by the word “som” (which in this case was introducing a
Hyphenations (divided words), deemed as important for the dyslexic reader. A hyphenation is a word with a hyphen in between two syllables. In this work hyphenations over two lines have been investigated. They begin on one line, and continue on the next line, with a hyphen after the first part of the word (See figure 1). Both the hard texts had a total of three hyphenations each.

Numbers written out as letters, which is not deemed as important for the dyslexic reader. Examples is writing “one” instead of “1”. The easy text had 2 examples of writing numbers as letters. However Luz Rello et al (2013) showed that the mean fixation time for a dyslexic group increased when writing the number as letters, something that did not affect their control group (Rello et.AI, 2013).

The last experimental manipulation was creating a stepping stone to the text, by writing the first few (3-5) words in four paragraphs in bold style, such as discussed on the workshop mentioned earlier (Dyslexiförbundet, 2018)

Figure 1, showing a divided word (marked in red).

The nonexperimental manipulations were:

The texts were all split into two columns. Both columns were narrower than a total of 60 letters wide, according to guidelines from the Swedish dyslexia association. Columns width were found not to have an impact on the reading performance in a study made by Rello and Baeza-Yates (2015), but that personal preferences agreed with the guidelines, that the columns should be less than 66 characters wide (60 according to the guidelines) (Rello & Baeza-Yates, 2015).

The font selected was Calibri, a sans-serif font selected based on recommendations for making the text easily available to a dyslexic reader (Rello & Baeza-Yates, 2013)

Since the texts were collected from the national tests they should be of comparable difficulty, however the web application TeCST was used to generate a second opinion on whether or not the stimuli were comparable (and therefore able to merge the data from the “easy” text 1 and “easy” text 2), and get a relatively objective measurement of the texts readability, (see results, table 3 for details) as opposed to the coded and subjective opinion of the participants.

Falkenjack, Rennes, Fahlborg, Johansson and Jönsson (2017) writes that TeCST is a tool to analyze and help simplify texts, and is an application consisting of several parts. One of which is SCREAM. They explain that SCREAM (Swedish Compound REAdablitiy Metric) consisted of 117 different text complexity measures at the time of writing, of which TeCST uses only a subset (Falkenjack et al., 2017).

3.3 Procedure

The participant gave verbal informed consent after being instructed that the experiment was voluntary, they were allowed to quit the experiment whenever without having to name a reason, and that the data would be handled with care such that their identities and their individual performances would not be able to be traced back to them. The information was provided both verbally and in written form.

After agreeing to the conditions they got to sit down at between 60-70 cm from the screen, and a calibration of the eye tracking equipment took place. The calibration was a five point
calibration, with validation as a quality check. The stimuli were showed in a balanced manner with 4 different groups (see table 2). These different groups were designed to counter any form of order effects.

Table 2, the order of the texts, as seen groups 1 and 4 read the same texts with the same conditions, however in a countered order. The same applies to groups 2 and 3.

<table>
<thead>
<tr>
<th></th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Group 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>First text:</td>
<td>Easy Text 1</td>
<td>Hard text 1</td>
<td>Easy text 2</td>
<td>Hard text 2</td>
</tr>
<tr>
<td>Second text:</td>
<td>Hard text 2</td>
<td>Easy Text 2</td>
<td>Hard text 1</td>
<td>Easy Text 1</td>
</tr>
</tbody>
</table>

Each participant got to read a short example text and answer an example question based on that text. This was made to make sure that the participants felt they could read naturally, even though the experiment leader were in the same room and they were placed in front of a camera recording their eye movements. After which written instructions told them that when they were ready the first text would be visible and that they were to take as much time as needed and to read it as they would naturally. It also informed them that there would be questions regarding the text they had read. After having read the text and answered the questions there was an instructing sentence showed prompting the participant to continue with the second text when ready. The participant set the pace by being allowed to spend as much time as they wanted reading the texts, but were not however allowed to review the text after being presented with the questions. The given task affects the way one looks at a stimuli, fixation time for example can vary depending on which instructions the participant has been given. (Greene, Liu & Wolfe, 2012). Hence the participants needed to read the entirety of the text carefully to be able to answer the questions presented. If the questions would have been showed before the text the reading pattern would have looked quite different, and the task would have been transformed to scanning the text rather than reading the text in its entirety.

3.4 Design

To examine the results within the groups, paired-samples t-tests were conducted. Effect sizes was calculated using Cohens D (Field, 2013). To see if a condition was harder for one group than the other the effect sizes was compared. (Mean 2 – Mean 1 / SD1, where 1 was the easy text and 2 the hard text, see Equation 1).

For the data that proved not to fulfil the assumptions for the t-test (not being normally distributed), Wilcoxon Signed Rank test was conducted (Field, 2013).

3.5 Areas Of Interests

Areas of interest (AOI) were defined around the manipulations in the texts. They were defined in such a way so that if words were written in bold in the “easy” condition the same words were covered in an AOI in the “hard” text, to be able to make comparisons between the two. The sizes of the different AOIs are shown in table 3.
Table 3, AOI sizes and their names. The leftmost column show which text the manipulation was made upon. The two next show the “easy” texts while the two last columns show the “hard” texts. The size of the AOI is measured in pixels.

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>AOI Name Easy</th>
<th>AOI Size Easy</th>
<th>AOI Size Hard</th>
<th>AOI Name Hard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text 1</td>
<td>bold style 1</td>
<td>6160</td>
<td>4788</td>
<td>non-bold 1</td>
</tr>
<tr>
<td>Text 1</td>
<td>bold style 2</td>
<td>5880</td>
<td>5974</td>
<td>non-bold 2</td>
</tr>
<tr>
<td>Text 1</td>
<td>bold style 3</td>
<td>4300</td>
<td>3204</td>
<td>non-bold 3</td>
</tr>
<tr>
<td>Text 1</td>
<td>bold style 4</td>
<td>5292</td>
<td>4680</td>
<td>non-bold 4</td>
</tr>
<tr>
<td>Text 1</td>
<td>non-divided word 1</td>
<td>2875</td>
<td>2904</td>
<td>divided word 1</td>
</tr>
<tr>
<td>Text 1</td>
<td>non-divided word 2</td>
<td>2880</td>
<td>2723</td>
<td>divided word 2</td>
</tr>
<tr>
<td>Text 1</td>
<td>non-divided word 3</td>
<td>2912</td>
<td>2862</td>
<td>divided word 3</td>
</tr>
<tr>
<td>Text 1</td>
<td>short sentence 1</td>
<td>40392</td>
<td>40280</td>
<td>long sentence 1</td>
</tr>
<tr>
<td>Text 1</td>
<td>short sentence 2</td>
<td>34146</td>
<td>35112</td>
<td>long sentence 2</td>
</tr>
<tr>
<td>Text 1</td>
<td>number 1</td>
<td>1704</td>
<td>696</td>
<td>number 1</td>
</tr>
<tr>
<td>Text 1</td>
<td>number 2</td>
<td>1863</td>
<td>1404</td>
<td>number 2</td>
</tr>
<tr>
<td>Text 2</td>
<td>bold style 1</td>
<td>4080</td>
<td>3588</td>
<td>non-bold 1</td>
</tr>
<tr>
<td>Text 2</td>
<td>bold style 2</td>
<td>2600</td>
<td>2496</td>
<td>non-bold 2</td>
</tr>
<tr>
<td>Text 2</td>
<td>bold style 3</td>
<td>15180</td>
<td>15930</td>
<td>non-bold 3</td>
</tr>
<tr>
<td>Text 2</td>
<td>bold style 4</td>
<td>4770</td>
<td>4256</td>
<td>non-bold 4</td>
</tr>
<tr>
<td>Text 2</td>
<td>non-divided word 1</td>
<td>1898</td>
<td>2442</td>
<td>divided word 1</td>
</tr>
<tr>
<td>Text 2</td>
<td>non-divided word 2</td>
<td>3024</td>
<td>3156</td>
<td>divided word 2</td>
</tr>
<tr>
<td>Text 2</td>
<td>non-divided word 3</td>
<td>1988</td>
<td>2338</td>
<td>divided word 3</td>
</tr>
<tr>
<td>Text 2</td>
<td>short sentence 1</td>
<td>27968</td>
<td>32488</td>
<td>long sentence 1</td>
</tr>
<tr>
<td>Text 2</td>
<td>short sentence 2</td>
<td>24147</td>
<td>23563</td>
<td>long sentence 2</td>
</tr>
<tr>
<td>Text 2</td>
<td>number 1</td>
<td>1060</td>
<td>918</td>
<td>number 1</td>
</tr>
<tr>
<td>Text 2</td>
<td>number 2</td>
<td>1960</td>
<td>736</td>
<td>number 2</td>
</tr>
</tbody>
</table>

As can be seen in table 3 the AOI size differed slightly between the “easy” and the “hard” versions of the text. For an example of this see figure 2 and figure 3 that shows the AOI “short sentence 1” in the easy and hard version of text 1.

**Figure 2. The AOI covering short sentence 1 in the “easy” version of text 1.**
Measurements

The data were of two categories, absolute and coded. The absolute data consisted of fixation time (measured in ms.) and the number of fixations. Fixation time is the time spent on fixating on a single point (SMI, n.d). These were measured in each AOI and then normalized: for each participant the time spent focusing on a certain AOI was divided on the total time that the participant spent reading that text. The same was done with the fixations. By dealing with the proportion of the time rather than the absolute value of the time itself comparisons are able to be made between the groups and see a meaningful analysis instead of comparing the speed with which the different groups read (Dyslexiföreningen, n.d).

Additionally data were gathered that were of a more qualitative nature: the reading comprehension was tested with two questions regarding the text itself, and the readability of the text was measured with a 5 point Likert scale, ranging from 1: hard to read, 3: not hard, nor easy to read and 5: easy to read.

TeCST values were deemed similar enough (see table 4) for the data points from the “easy” texts to be combined and therefore get a more valid result, since the number of data collection points double in number, (see table 5). The same was done with the “hard” texts.

Table 4 shows the scores for a number of different measures of readability, both TeCST. Falkenjack et al (2017) summarizes the different measurements that TeCST generates while analyzing a text. The different measurements were:

- The mean word length, calculated as the average number of characters per word (Falkenjack et al., 2017).
- The mean sentence length, average number of words per sentence (Falkenjack et al., 2017).
- Lix: readability index (Läsbarhetsindex): based on categorical word frequencies in the Swedish vocabulary “SweVoc”. Shows the ratio of words longer than 6 characters, together with average number of words in a sentence. (Falkenjack et al., 2017)
- Ovix: word variation index (Ordvariationsindex): The quota of unique words and the total number of words in the text. (TeCST, n.d)
- SweVocTotal: The number of unique words from the vocabulary of easy Swedish words, SweVoc, in the sentence. (Falkenjack et al. 2017)
- Mean dependency distance: The average distance between the dependant and the head for each dependency. (TeCST, n.d)
Table 4, number of data collection points for both groups

<table>
<thead>
<tr>
<th></th>
<th>Bold/Non-bold</th>
<th>Divided/non-divided words</th>
<th>Long/short sentences</th>
<th>Numbers as letters/numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dyslexic group</td>
<td>32</td>
<td>24</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>Control group</td>
<td>36</td>
<td>27</td>
<td>18</td>
<td>18</td>
</tr>
</tbody>
</table>

The coded measure were the answers to the reading comprehension questions, in percent correct. The available results possible were 0, 25, 50, 75 and 100 % correct. Another subjective measure were the 5 graded scale on which the participants got to answer how hard to read (1), “neither hard nor easy to read” (3) or easy to read (5).

3.7 Readability

The texts were analyzed by TeCST and the results in table 5 were given.

Table 5, TeCST scores for both the texts

<table>
<thead>
<tr>
<th>Measurement</th>
<th>Text 1 Easy</th>
<th>Text 2 Easy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean word length</td>
<td>4,4</td>
<td>4,7</td>
</tr>
<tr>
<td>Mean sentence length</td>
<td>16,48 ± 7,82</td>
<td>15,43 ± 8,58</td>
</tr>
<tr>
<td>Lix: Readability index</td>
<td>37,423</td>
<td>39,407</td>
</tr>
<tr>
<td>Ovix: word variation index</td>
<td>59,015</td>
<td>59,627</td>
</tr>
<tr>
<td>SweVocTotal</td>
<td>54,01%</td>
<td>52,92%</td>
</tr>
<tr>
<td>Mean dependency length</td>
<td>2,45</td>
<td>2,44</td>
</tr>
</tbody>
</table>
4. Results

This chapter presents the result of the readability analysis given the amount of correct answers, and also the statistical analysis.

4.1 Perceived readability

The mean of the grades of readability and the percentage of correct answers corresponding to each text is shown in Table 6. The values were imported into Microsoft Excel and the mean was calculated for the readability of both texts, and the mean of correct answers that were given.

Table 6, mean values of readability and percent correct answers.

<table>
<thead>
<tr>
<th></th>
<th>Readability Easy text</th>
<th>% correct Easy text</th>
<th>Readability Hard text</th>
<th>% correct Hard text</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dyslexia</td>
<td>3,429</td>
<td>0.714</td>
<td>3,778</td>
<td>0.944</td>
</tr>
<tr>
<td>Control</td>
<td>3,571</td>
<td>0.643</td>
<td>3,667</td>
<td>0.833</td>
</tr>
</tbody>
</table>

4.2 Statistics

All statistical analysis except for calculating the effect size was done in IBM SPSS. The effect size was calculated in Microsoft Excel according to the formula: Mean1 minus Mean2 divided by the standard deviation for Mean 2 (see figure 1), where the standard deviation was taken from the easy text (Field, 2013).

Equation 1, the formula for calculating Cohens d within the group according to Field (2013)

\[
\text{Cohens } d = \frac{M_1 - M_2}{s_2}
\]

The data was tested for normality, using SPSS. Both Kolmogorov-Smirnov tests and Shapiro-Wilk tests were performed simultaneously. Normality can be shown graphically via histograms (Field, 2013) but also via Q-Q plots (Laerd, n.d). Examining the data graphically can have its advantages however unless the examiner has a lot of experience in doing so, the numerical methods might be more safe (Laerd, n.d). The Shapiro-Wilk test is more suitable for small sample sizes, and thus that test was chosen (Laerd, n.d).

To compare the means of the different group’s paired-samples t-tests were done on the data, both for the control group and for the dyslexic group.

The Shapiro-Wilk test revealed that more than half the data was not normally distributed. Wilcoxon Signed Rank test was performed on the data that was not normal. Effect size was calculated on the Wilcoxon Signed Rank test results by equation 2, according to Field (2013):

Equation 2. The formula for calculating effect size r for Wilcoxon Signed Rank test

\[
r = \frac{z}{\sqrt{N}}
\]
Table 7. Results from statistical analysis for the dyslexic group. As described above the values were calculated by running paired-samples t-tests, and effect size was calculated using Cohens d for the normally distributed data. The rest were calculated with Wilcoxon Signed Rank test and r as effect size. Note that the t-tests are reported with a t, and the non-parametric tests are reported with a z.

<table>
<thead>
<tr>
<th>Dyslexic group statistics</th>
<th>Effect size:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bold style:</td>
<td></td>
</tr>
<tr>
<td>Fix.Time: z = 0,355</td>
<td>p = 0,722</td>
</tr>
<tr>
<td>Fixations: t(31) = 0,748</td>
<td>p = 0,460</td>
</tr>
<tr>
<td>Divided words:</td>
<td></td>
</tr>
<tr>
<td>Fix.time: z = 0,457</td>
<td>p = 0,648</td>
</tr>
<tr>
<td>Fixations: z = -0,371</td>
<td>p = 0,710</td>
</tr>
<tr>
<td>Sentences:</td>
<td></td>
</tr>
<tr>
<td>Fix.time: t(15) = -0,811</td>
<td>p = 0,430</td>
</tr>
<tr>
<td>Fixations: t(15) = -1,118</td>
<td>p = 0,281</td>
</tr>
<tr>
<td>Numbers:</td>
<td></td>
</tr>
<tr>
<td>Fix.time: z = 1,161</td>
<td>p = 0,245</td>
</tr>
<tr>
<td>Fixations: z = 0,345</td>
<td>p = 0,730</td>
</tr>
</tbody>
</table>

Table 8, results from statistical analysis for the control group. The values are reported and calculated in the same way as described above and in table 7.

<table>
<thead>
<tr>
<th>Control group statistical results:</th>
<th>Effect size:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bold style:</td>
<td></td>
</tr>
<tr>
<td>Fix.time: z = 1,005</td>
<td>p = 0,315</td>
</tr>
<tr>
<td>Fixations: t(35) = -0,380</td>
<td>p = 0,706</td>
</tr>
<tr>
<td>Divided words:</td>
<td></td>
</tr>
<tr>
<td>Fix.time: z = -2,248</td>
<td>p = 0,025</td>
</tr>
<tr>
<td>Fixations: z = -1,207</td>
<td>p = 0,228</td>
</tr>
<tr>
<td>Sentences:</td>
<td></td>
</tr>
<tr>
<td>Fix.time: z = 1,546</td>
<td>p = 0,122</td>
</tr>
<tr>
<td>Fixations: z = 0,675</td>
<td>p = 0,500</td>
</tr>
<tr>
<td>Numbers:</td>
<td></td>
</tr>
<tr>
<td>Fix.time: z = -1,59</td>
<td>p = 0,112</td>
</tr>
<tr>
<td>Fixations: z = -1,931</td>
<td>p = 0,053</td>
</tr>
</tbody>
</table>

As seen in table 7 and 8, the only significant value was found when measuring fixation time for divided words for the control group (z = -2,248, p = 0,025, with an effect size of -0,433).

Kendalls Tau b was calculated to see if there was a correlation between the perceived readability of the texts and the comprehension as tested by the questions presented after the text was read. The relationship between time spent reading and the perceived readability was also investigated. Kendalls Tau b was chosen since it fits the best with small sample sizes (Field, 2013).

None of the results were even close to be significant, with p-values between 0,115 and 0,794 and correlation coefficients were between -0,060 for readability of the hard text and the correct answers, while the highest value of the coefficient was 0,363 for readability for the easy texts and the correct answers for that text.
5. Discussion

5.1 About the results

The only result that was significant in the experiment was found when examining divided words. Upon closer inspection of the data the mean fixation time of words spelled together and words split up over two rows, it was found that the time spent fixating on the AOI containing the split up word was smaller than the time fixating on the word spelled together. This could have several explanations. The reader might be able to predict what the next part of the word that has been divided might be, based on context and the first part of the word. Otherwise the reader might have developed a reading strategy and bypassed the part of the AOI that covered the second part of the divided word. Since the eye tracking is measuring where your gaze is focused, it might have been outside of the AOI, but still close enough to the actual word to process it just by having it in the periphery of the gaze, hence leading to shorter fixation time ON the actual word.

The same can be said about the numbers, where the number of fixations were not significant, but had an effect size of -0.455, which is considered medium sized (see above). The difference in AOI size, (something that is quite natural. Say for example if one were to compare the size of “twenty” with the size of “20”) might have affected this result. The interesting part of this is that Rello et al (2013) found that written with digits were significantly more readable than numbers written spelled out in the dyslexic group. This is something that the Persona developed by the Swedish dyslexia association claimed not to be important. The findings in this study, however not significant, also points towards the use of digits when writing numbers to decrease the fixation time needed, hence increasing the readability of the document.

The effect sizes were rather sizeable, so if more participants had taken part and the p-value had been significant, the effect size of the tests would have been of a fair size. Paradoxically, both the control group and the dyslexic group judged the “hard” texts to be more readable than the “easy” one as seen in table 6. The mean percentage of correct answers were also higher for the “hard” texts. This could have several reasons, one explanation is that the low number of participants, see “about the method” below.

5.2 About the method

The number of participants are low, however since there are several data collection points for each participant some credibility can still be found in the findings. Since the sample size is so limited however, the results should be seen as tendencies and patterns rather than definite answers to if the persona that is being evaluated is correct or not. Another reason the tests did not say much might have been because the manipulations made are rather small. When designing the stimuli texts efforts were made to keep the ecological validity high. Therefore the number of experimental manipulations were limited, see “equipment” for more details. The reasoning behind this was the ethical aspect, the dyslexic group might have individuals who prefer not to read, and to simply do so in a lab setting might cause some to stress to those persons. If the texts were too hard to read and comprehend the participants might have decided to opt out half way through the experiment.

The texts were from the secondary school level (Swedish: “Gymnasial nivå”) and even though all participants had at least secondary school as the highest finished education level, they might have been too complex or filled with information that could prove relevant when the questions later were presented. The problem with presenting the questions before letting the participant read through the entirety of the text is stated above in “Procedure”.

21
In this work the participants were slightly skewed towards female participants, as seen in table 1. The individual variation could have a profound impact on the results because of the limited sample size, which is a reason for not stating anything with certainty when presenting the interpretation of the results.

The tracking ratio in one of the participants was low, (71.7 %). Some data from these participants might have been lost, but since the participant were dyslexic and given the problems in finding participants a decision was made and the data was not discarded. During the reading of the other text the participant had a tracking ratio of 97.1 % so it must have been a temporary problem with the tracking.

In the experiments done by Holsanova, Rahm and Holmqvist (2006) and Holsanova, Holmberg and Holmqvist (2009) the participants got to read newspapers freely, with a headtracking device. Such equipment was not available during this study, and as a consequence the participants needed to keep their heads relatively still. The eye tracking device used seemed to work best at a distance of roughly 60-70cm from the screen on which the stimuli was presented. This might have affected the reading since it might have been experienced as a distraction to sit still. During the recording of the data the participants were monitored however, as unobtrusive as possible, by the experiment leader sitting by another computer in the room. On that computer the distance between the screen and the participant was shown. On a few occurrences participants needed to be asked to pull their chair forward or backwards since they relaxed and slumped together a bit after the calibration event was over, causing them to be outside of the comfort zone of the device. This was done in between the reading events, careful not to interrupt the reading process.

According to one participant the font used looked odd, a total of three participants commented on the columns, and that they were having the opposite effect of what was intended. Rello and Baeza-Yates (2015) did an eye tracking study and found that columns did not have an effect for dyslexic readers, but between 44 and 66 characters were preferred by their participants.

However the columns in this study was 42 monospaced characters wide, so perhaps it was the distance in between the columns that was the issue for the participants who raised their opinion.

5.3 Future research

Further research should aim towards conducting larger scale studies than this thesis work has done. To be able to correctly answer the question whether the guidelines improves readability or not more participants will be needed, further work will be needed to be spent on the perfecting the stimuli, and perhaps even investigate the effect that pictures and other non-text elements can have on reading, such as Holsanova, Rahm and Holmqvist did 2006 and Holsanova, Holmberg and Holmqvist did 2008 in newspaper spreads. The application TeCST could also be evaluated, since in this work only the analysis part of it was used, where there are also tools made for improving the readability by simplifying, and to summarize a text.
6. Conclusion

Given the difficulties finding participants, the results are not certain. A difference in readability was not found with the used method and equipment, however the low number of the sample size must be stressed. The effect sizes ranged between small to a rather impressive size, so given a larger sample size the individual variation would not have mattered as much and the results would point more clearly towards a definite answer to the question “Does the guidelines affect the readability for people with dyslexia”. The answer found in this study is leaning towards a no. That there is no difference in readability between the easily read text and the harder text, but for a definite answer to the question of the reliability and validity of the guidelines, further studies could focus on a more detailed scale on which to measure readability than the one used in this study. That result could also be compared to the results given from tools such as TeCST.
7. References


TeCST.  *FAQ* retrieved May 30, 2018 from https://www.ida.liu.se/projects/scream/webapp/


Vetenskapsrådet (2008). Kort om dyslexi [Broschure]

Workshop with Dyslexiförbundet (March 21, 2018)