Noname manuscript No.
(will be inserted by the editor)

The Impact of Misspellings on Reading Comprehension for People with Dyslexia

Ricardo Baeza-Yates · Luz Rello

Received: date / Accepted: date

Abstract We present two user studies with an overall number of 122 participants (61 with dyslexia) that explore the impact of the presence of misspellings on reading performance. Participants were asked to read a set of texts: without misspellings, with different rates of misspellings, as well as texts with jumbled letters. Readability was measured via eye tracking measures and comprehension via questionnaires. The main conclusion is that the presence of errors in the text does not impact the reading performance of people with dyslexia as much as it does for people without dyslexia.

Keywords Errors · Misspellings · Eye tracking · Readability · Comprehension · Dyslexia

1 Description for Program

We present two user studies with an overall number of 122 participants (61 with dyslexia) that explore the impact of the presence of misspellings on reading performance. Participants were asked to read a set of texts: without misspellings, with different rates of misspellings, as well as texts with jumbled letters. Readability was measured via eye tracking measures and comprehension via questionnaires. The main conclusion is that the presence of errors in the text does not impact the reading performance of people with dyslexia as much as it does for people without dyslexia.

This paper was partially funded by a NIDRR grant, number H133A130057, from the Department of Education of the USA.

Universitat Pompeu Fabra
E-mail: rbaeza@acm.org

HCI Institute, Carnegie Mellon University
E-mail: luzrello@cs.cmu.edu

2 Purpose, Rationale and Citations

One of the reasons why humans can read is because we can recognize words. However, one of the key problems of dyslexia is inaccurate and slow word-recognition. The difficulties in word-recognition are not only present in children with dyslexia, also in dyslexic adults. They strongly relay on the use of spelling-sound information, syllabic information, and context for word recognition [4]. However, the exploration of word-recognition skills of people with dyslexia has been done using either texts with correct words o either non-words. Then, how do people with dyslexia read errors? At the same time people with dyslexia present poor spelling. How they recognize their own errors? How does the presence of errors affect the reading performance and comprehension of people with dyslexia? To which point can be the original word modified until is understood? Here we present the first study that measures the impact of the presence of dyslexic errors and text with jumbled letters in the readability and the comprehension of people with and without dyslexia using eye tracking.

Several researchers have explored how text modification affects word recognition. First, Rawlinson [23] in 1976 explored the impact of letter position in word recognition. One of the most relevant results of their study was that middle letter identification is independent of their position. In other words, letter randomization in the middle of the word had little effect on the ability of skilled readers to understand printed text [22]. Later in 2003, a text started to circulate in the Internet named as *the Cambridge University effect*,[1] even though the text was not related to any true research project that was ever conducted at the University of Cambridge. Others studies [1,7,37,12] have shown consistently that transposed-letter pseudo–words are perceptually very similar to their base words, and they tend to be (initially) misperceived as their corresponding base words. The comprehension of the participants was still high with transposed-letter texts in English [28] and Japanese [15], but the word recognition performance decreased in Hebrew [36]. To recognize words, people rely on the use of spelling-sound information, syllabic information, and context for word recognition. These difficulties in word-recognition are similar in both adults and dyslexic children [4]. In relation to word recognition it has been described the non-word reading deficit for English [19] and later for other languages with more straightforward grapheme-phone correspondences as German [39], Spanish [34] or Italian [35]. In these studies, non-words, that is, non-existing but possible words in language for instance *molsmi* or *brigbert* would be examples of difficult non-words, and *blem* or *tig* examples of easy non-words [19]. People with dyslexia presented more difficulties that the non-dyslexic population in both cases, non-word reading and word recognition. To the best of our knowledge the transposed-letter effect has not been studied in relationship with dyslexia, and the exploration of word-recognition skills of

[1] "Aoccdrnig to rsaceerh at Cmabrigde Uinervtisy, it deosn't mttaer in waht oredr the ltteers in a wrod are, the olny iprmoetnt tihng is taht the frist and lsat ltteer be at the rghit pclae. The rset can be a total mses and you can sitll raed it wouhit porbelm. Tihs is bcuseae the huamn mnid deos not raed ervey lteter by istlef, but the wrod as a wlohe."

people with dyslexia has been done using either texts with correct words o either non-words.

In this work, we explore the effect of orthographic dyslexic errors and letter-transposition on the readability and comprehension of people with and without dyslexia. To do so we conducted two experiments with a total of 122 participants (61 with dyslexia) using eye tracking to measure readability, comprehension, and error awareness of the participants. Our results show that in most of the cases the presence of errors in the text does not impact the reading performance (readability and comprehension) of people with dyslexia, while it does for people without dyslexia.

3 Methodology

To study the effect of written errors on text readability and comprehensibility on screen, we conducted two experiments with Spanish native speakers: *Errors I* and *Errors II*, with 44 participants (22 with dyslexia) and 78 participants (39 with dyslexia), respectively. All the participants had to read a set of texts with varying rates of written errors. Readability and comprehensibility were analyzed via eye tracking and comprehension tests. Via questionnaires we gathered the participants' subjective ratings of readability and comprehensibility, as well as their error awareness.

3.1 Ethics Statement

This study follow the ethics protocol of Universitat Pompeu Fabra for user studies. Volunteers who were not minors provided written informed consent, while those who were minors provided written assent, with written consent additionally obtained from their parents.

3.2 Design

We conducted two experiments over time first *Errors I* and second *Errors II*.

- **Errors I:** In our experimental design, [±Errors] served as an *independent variable* with 3 levels: [No errors], [16% errors−Explicit] and [16% errors+Explicit]. The condition [No errors] denotes that the text was presented without any errors, while [16% errors−Explicit] and [16% errors+Explicit] denote the conditions when the texts were presented with 12 errors per text (each text had 77 words, 16% error rate) and the participants knew that the text had errors. [16% errors−Explicit] denotes the condition where the text was presented with errors and the participant was not informed about the presence of errors. The condition [16% errors+Explicit] denotes when the participant knew in advance that the text to read had errors. Refer to Table 1 for a summary of the first *Errors I* experiment.

– In the second experiment **Errors II**, [±Errors] served as an *independent variable* with 4 levels: [8% errors−Explicit], [8% errors+Explicit], [50% errors short+Explicit] and [50% errors long+Explicit]. The condition [8% errors−Explicit] denotes when the text was presented with 6 errors (in a text of 75 words, 8% error rate) and the participant were not informed about the presence of errors in the text. The condition [8% errors+Explicit] denotes when the text was presented with 6 errors and the participant was explicitly informed about the presence of errors in the text. The conditions [50% errors short+Explicit] and [50% errors long+Explicit] denote when 49.3% of the words in the text (37 of 75 words per text) were presented with their letters in random other, except the first and last letter of the word. For instance, **pitraa (pirata, 'pirate')* or **mdnuo (mundo, 'world')*. These two last conditions are visually explicit as almost all words have errors. See Table 2 for a summary of the *Errors II* experiment.

We used a within-subject design, that is, each participant read all the texts, contributing to each of the conditions. To avoid sequence effects, we counterbalanced texts as far as possible. We needed to maintain that the [−Explicit] condition was always before that its equivalent in errors [+Explicit] condition, that is, [8% errors−Explicit] was always before [8% errors+Explicit] and [16% errors−Explicit] was always before [16% errors+Explicit].

3.2.1 Dependent Variables

For quantifying objective *readability*, we used the following objective measures: *Reading Time*, *Fixation Duration*, and *Number of Fixations*. These data is extracted directly from the eye tracker software. For quantifying objective *comprehensibility*, we used a *Comprehension Score* as a dependent variable. To measure the readers subjective perception towards the text and the misspells we used *Readability Rating*, *Comprehensibility Rating* and *Error Awareness*.

- *Reading Time:* Shorter reading durations are preferred to longer ones since faster reading is related to more readable texts [38]. Therefore, we use *Reading Time*, that is, the time it takes a participant to completely read one text, as a measure of readability.
- *Fixation Duration:* When reading a text, the eye does not move contiguously over text, but alternates saccades and visual fixations, that is, jumps in short steps and rests over pieces of text. *Fixation Duration* denotes how long the eye rests still on a single spot of the text. We use the average of the fixation durations as a metric for readability.
- *Number of Fixations:* We use the total number of fixations while reading a text. *Fixation Duration* and *Number of Fixations* have been shown to be a valid indicator of readability, as eye movement measures can be used to infer moment-to-moment cognitive processes in reading. Shorter fixations are associated with better readability while a greater number of fixations and longer fixations can indicate that processing loads are greater [25]. For

instance, people without dyslexia present more and longer fixations on low frequency words than on high frequency words [11,13,21,26,27].
There are four studies that show why fixation duration and number of fixations are valid indicators for people with dyslexia. These are the studies by Hyöna and Olson [10], Pirozzolo and Rayner [18], Olson *et al.* [16] and Rayner [24].

- *Comprehension Score:* Moving the eyes over a text does not guarantee its comprehension. To check that the text was not only read, but also understood, we measure text comprehension using questionnaires. We compute the *Comprehension Score* as the average of the answers.
- *Readability Rating:* We use this rating for quantifying subjective *readability*. The participants rated on a five-point Likert scale, how easy was to read the text.
- *Comprehensibility Rating:* We use this rating for quantifying subjective *comprehensibility*. The participants rated, on a five-point Likert scale, how easy was to understand the text.
- *Error Awareness:* To measure the awareness of the presence of errors in the text we defined *Error Awareness* as how the participant perceived the number of errors in the texts.

3.3 Participants

Overall, 122 participants undertook the two experiments. For *Errors I* we had 22 Spanish speakers (12 female, 10 male) with a confirmed diagnosis of dyslexia (Group D) and 22 Spanish speakers (13 female, 9 male) without dyslexia, which served as a control group (Group N). For *Errors II* we had 39 Spanish speakers (20 female, 19 male) with a confirmed diagnosis of dyslexia (Group D) and 39 Spanish speakers (23 female, 16 male) without dyslexia, which served as a control group (Group N, non-dyslexia). The participants were asked about the languages they speak, their level of studies and about their reading habits (details are given in Tables 1 and 2.).

All the participants had normal or corrected to normal vision. None of the participants were screened for visual stress (Meares-Irlen) syndrome. Participants from Group D were asked to bring their diagnoses to the experiment, to guarantee that dyslexia was diagnosed in an authorized centre or hospital.

3.4 Materials

To isolate the effects of misspellings on the readability and comprehension of people with and without dyslexia we created the following materials.

3.4.1 Texts and Misspellings

All the texts used in the experiments meet the comparability requirements because they all share the parameters commonly used to compute text com-

Errors I		
Design	Within-subjects	
Independent Variables	[±Errors]	[No errors] [16% errors−Explicit] [16% errors+Explicit]
	Error Awareness	[+Explicit] [−Explicit]
Dependent Variables	*Reading Time* *Fixation Duration* *Number of Fixations*	(objective readability)
	Comprehension Score	(objective comprehensibility)
	Error Awareness Rate	(subjective error awareness)
Participants	Group D (22 participants)	12 female, 10 male **Age:** range from 13 to 37 ($\bar{x}$ = 20.59, s = 8.32) **Bilingual:** Catalan (12), English (1), Italian (1) **Education:** university (8), high school (11), no higher education (3) **Reading:** more than 8 hours (3), 4-8 hours (8), less than 4 hours (11) per day
	Group N (22 participants)	13 female, 9 male **Age:** range from 13 to 35 ($\bar{x}$ = 21.27, s = 8.89) **Bilingual:** Catalan (14), English (2), French (1) **Education:** university (12), high school (9), no higher education (1) **Reading:** more than 8 hours (0), 4-8 hours (9), less than 4 hours (13) per day
Materials	Base Texts	3 texts
	Errors	24 errors (12 errors per text)
	Text Presentation	
	Comprehension Questionnaire	3 inferential and 3 literal items (2 items of each type per text)
	Error Survey	2 items (1 item/text with errors)
Equipment	Eye tracker Tobii 1750	
Procedure	**Steps:** Instructions, demographic questionnaire, reading task (×3) comprehension questionnaires (×3) and error survey (×2)	

Table 1 Methodological summary of the first experiment *Errors I*.

plexity [6]. The base texts (a) had the same genre; (b) the same style; (c) were about and similar topics: news about culture; (d) had very similar number of words ranging from 75 to 77 words. The texts used for [16% errors±Explicit] and [8% errors±Explicit] conditions were the same in both experiments and had word length average of $\bar{x} = 4.83 \pm 3.07$. The words in the text [50% errors short+Explicit] were shorter, with a word length average of $\bar{x} = 4.27 \pm 2.39$. The words in the text [50% errors long+Explicit] were longer, with a word length average of $\bar{x} = 5.12 \pm 3.65$.

The misspellings used in conditions [8% errors−Explicit], [8% errors+Explicit], [16% errors−Explicit] and [16% errors+Explicit] are errors written by people with dyslexia extracted from *DysList*, a list of 1,171 different word-error pairs extracted from 83 texts written by children with dyslexia [32]. We used a total of 24 dyslexic errors *(Errors I* used the same texts and errors –half less– than *Errors II)*. See the misspellings used in the Appendix.

Errors II		
Design	Within-subjects	
Independent Variables	[±Errors]	[8% errors−Explicit] [8% errors+Explicit] [50% errors short+Explicit] [50% errors long+Explicit]
	Error Awareness	[+Explicit] [−Explicit]
Dependent Variables	*Reading Time* *Fixation Duration* *Number of Fixations*	(objective readability)
	Comprehension Score	(objective comprehensibility)
	Readability Rating	(subjective readability)
	Comprehensibility Rating	(subjective comprehensibility)
	Error Awareness Rate	(subjective error awareness)
Participants	Group D (39 participants)	20 female, 19 male **Age:** range from 11 to 45 ($\bar{x}$ = 21.15, s = 9.39) **Bilingual:** Catalan (13), English (1), French (1) **Education:** university (16), high school (20), no higher education (3) **Reading:** more than 5 hours (1), 3-5 hours (11), less than 2 hours (27) per day
	Group N (39 participants)	23 female, 16 male **Age:** range from 11 to 43 ($\bar{x}$ = 26.56, s = 8,79) **Bilingual:** Catalan (11), English (3) **Education:** university (25), high school (12), no higher education (2) **Reading:** more than 5 hours (4), 3-5 hours (17), less than 2 hours (18) per day
Materials	Base Texts	4 texts
	Errors	12 errors (6 errors per text) 50% words with random letters/text
	Text Presentation	
	Comprehension Quest.	4 inferential and 4 literal items (2 items of each type/text)
	Sub. Readability Quest.	4 Likert scales
	Sub. Comprehension Quest.	4 Likert scales
	Error Survey	4 items (1 item/text with errors)
Equipment	Eye tracker Tobii 1750	
Procedure	**Steps:** Instructions, demographic questionnaire, reading task (×4) comprehension questionnaires (×4), subjective readability questionnaire (×4) subjective comprehension questionnaire (×4) and error survey (×4)	

Table 2 Methodological summary of the second experiment *Errors II*.

3.4.2 Questionnaires

Comprehension Questions: We used two comprehension items (one inferential and one literal) for each of the texts. Inferential questions require a deep understanding of the text. Literal questions can be answered directly from the text. We included these items in multiple-choice questions with three possible choices. The order of the correct answer was counterbalanced. The difficulty of the questions chosen was similar. One of the choices was always correct. The rest could be either wrong or partially correct. To compute the *Comprehension Score*, the choices counted 100%, 50%, and 0%, for the correct, partially correct and wrong answer, respectively.

¿Cuántos errores tenía el texto? *'How many errors had the text?:'*
- □ 100% de palabras. *'100% of misspelled words.'*
- □ 75% de palabras. *'75% of misspelled words.'*
- □ 50% de palabras. *'50% of misspelled words.'*
- □ 25% de palabras. *'25% of misspelled words.'*
- □ 0% de palabras. *'0% of misspelled words.'*

Fig. 1 Error survey item.

Subjective Readability and Comprehension Questionnaires: To quantify the *Readability Rating* and the *Comprehensibility Rating* we used questionnaires. For each of the conditions, the participants rated on a five-point Likert scale, to which extent the text was easy to read and understand.

Error Survey: After reading each of the texts, the participants were asked to estimate how many errors the texts had. In Figure 1 we show the survey items for condition [50%±explicit].

3.4.3 Text Presentation

We used the same layout for all the texts taking into account previous literature findings because the presentation of the text influences the reading performance of people with dyslexia [9]. The text was unjustified text since justified text alignment produces irregular spacing between words [2,17,20]. The font size used was 20 points [33] and the column width did not exceed 70 characters/column, as recommended by the British Dyslexia Association [3]. The color used was black font on crème background[2] [3,30].

We choose to present the texts in *Arial* because of three reasons. First, *Arial* is the most common font used on screen for the Web [5]. Second, *Arial* is highly recommended in literature. For instance, Evett and Brown [8] put in comparison recommendations for readers with low vision and dyslexia, and both groups agree in using *Arial* and *Comic Sans*. The British Dyslexia Association also recommends using *Arial*. Third, in Lockley's [14] study, *Arial* was the preferred font. Also in an experiment using eye tracking with 48 participants with dyslexia, *Arial* was the font that lead to significantly shorter reading time [29].

3.5 Procedure

The experiments were conducted at the Universitat Pompeu Fabra and lasted from 15 to 20 minutes each. They took place in a quiet room, where the participant was alone with the interviewer (second author of this paper), so that the participants could concentrate. Each participant performed the following

[2] The CYMK are crème (FAFAC8) and black (000000). Color difference: 700, Brightness difference: 244.

steps. After they signed the on-line consent, we began with a questionnaire that was designed to collect demographic information. They were asked to read the texts in silence and complete first the comprehension questionnaires, and then the error survey and the subjective ratings questionnaires. The eye tracker recorded their reading and they could not look back on the text when they had to answer the questions.

4 Results

In this section, we present the analyses of the data of both groups, D and C. Shapiro-Wilk tests showed that only some of the data sets were normally distributed. Also, Levene tests showed that the data sets were not homogeneous.

Hence, to study the effects of the conditions for readability and comprehensibility we used the two-way Friedman's non-parametric test for repeated measures plus a complete pairwise Wilcoxon rank sum post-hoc comparison test with a Bonferroni correction that includes the adjustment of the significance level. Then, to show effects of the conditions between groups D and C, we divided the data for each group and used Friedman's non-parametric test for repeated measures plus a complete pairwise Wilcoxon rank sum post-hoc comparison test with a Bonferroni adjustment.

Following we present the results for the readability, comprehension, reading efficiency (calculated using both readability and comprehension), the subjective readability and comprehensibility as well as error awareness.

4.1 Readability

In both experiments, there was a significant effect of [±Errors] on the three readability measures: (i) *Reading Time* ($\chi^2(2) = 20.61,\ p = 0.001$) for *Errors I* and $\chi^2(3) = 102.33,\ p < 0.001$) for *Errors II* (Figure 2); (ii) *Fixation Duration* ($\chi^2(2) = 20.61,\ p < 0.001$) for *Errors I* and ($\chi^2(3) = 54.87,\ p < 0.001$) for *Errors II* (Figure 3); and (iii) *Number of Fixations* ($\chi^2(2) = 8.72,\ p = 0.013$) and ($\chi^2(3) = 72.00,\ p < 0.001$) for both experiments respectively (Figure 4). The results of the post-hoc tests show that:

- **Between Groups:** In the first experiment, participants with dyslexia had significantly longer *Reading Times* ($\bar{x} = 28.52$, $s = 15.51$ seconds) than the participants without dyslexia ($\bar{x} = 17.69$, $s = 8.04$ seconds, $p < 0.001$). Group D had also significantly longer *Fixation Duration* ($\bar{x} = 0.23$, $s = 0.07$ seconds) than Group N ($\bar{x} = 0.21$, $s = 0.04$ seconds, $p < 0.001$). Consistently, participants with dyslexia had significantly more *Number of Fixations* ($\bar{x} = 117.48$, $s = 48.16$ fixations) than the participants without dyslexia ($\bar{x} = 82.86$, $s = 25.51$ fixations, $p < 0.001$).
  Consistently, in the second experiment *Errors II*, participants with dyslexia had significantly longer *Reading Times* ($\bar{x} = 56.98$, $s = 34.47$ seconds) than the participants without dyslexia ($\bar{x} = 36.50$, $s = 18.39$ seconds,

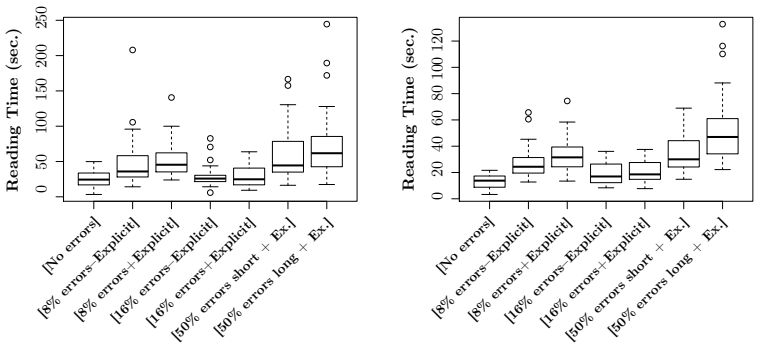

Fig. 2 *Reading Time* means for *Errors I* and *Errors II* experiments for Group D (left) and Group N (right).

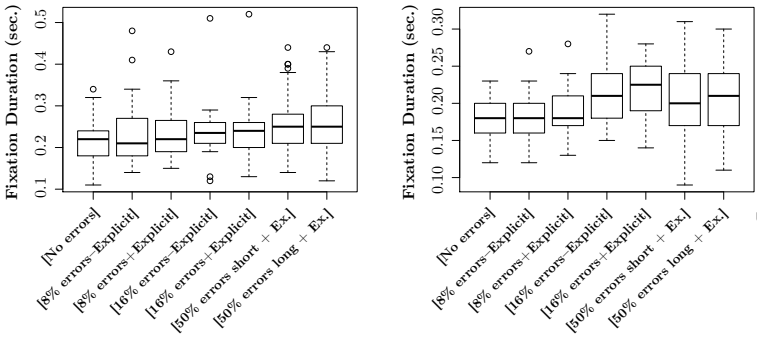

Fig. 3 *Fixation Duration* means for *Errors I* and *Errors II* experiments for Group D (left) and Group N (right).

$p < 0.001$). They had significantly longer *Fixation Duration* ($\bar{x} = 0.24$, $s = 0.07$ seconds) than Group N ($\bar{x} = 0.19$, $s = 0.04$ seconds, $p < 0.001$), and had significantly more *Number of Fixations* ($\bar{x} = 171.05$, $s = 82.53$) than the participants without dyslexia ($\bar{x} = 128.22$, $s = 63.70$, $p < 0.001$).

- **Group D:** In the first experiment *Errors I*, we did not find a significant effect of the conditions on *Reading Time*, *Fixation Duration* and *Number of Fixations* in Group D.
  However, in the second experiment *Errors II* there was a significant effect of [±Errors] on *Reading Time* and *Number of Fixations* in Group D. Texts with [8% errors−Explicit] lead to shorter reading times than texts with [50% errors long+Explicit] condition ($p < 0.001$). Likewise, texts with [8% errors−Explicit] lead to more *Number of Fixations* than texts with conditions [8% errors+Explicit] ($p = 0.052$), and [8% errors+Explicit] condition ($p = 0.005$).We did not find a significant effect of the conditions on *Fixation Duration* for Group D.
- **Group N:** In *Errors I*, there was a significant effect of [±Errors] on the three readability measures. Texts with [No errors] lead to shorter *Read-*

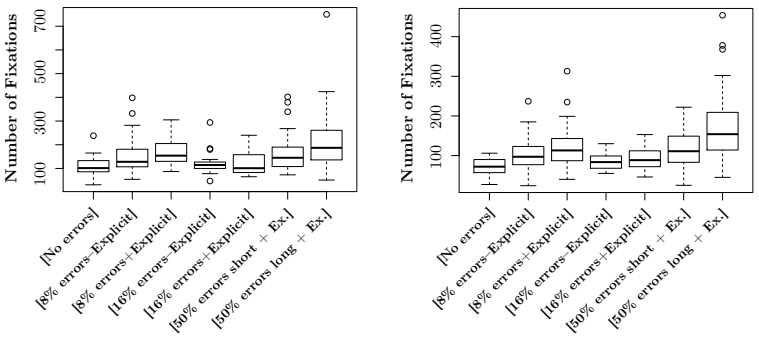

Fig. 4 *Number of Fixations* means for *Errors I* and *Errors II* experiments for Group D (left) and Group N (right).

ing Times than the text with [16% errors−Explicit] ($p = 0.001$) and with [16% errors+Explicit] ($p < 0.001$); they also lead to shorter *Fixation Durations* than the text with [16% errors−Explicit] ($p = 0.003$) and with [16% errors+Explicit] ($p < 0.001$); and finally texts with [No errors] presented fewer fixations than the text with [16% errors−Explicit] ($p = 0.011$) and with [16% errors+Explicit] ($p < 0.001$).
Evenly, there was a significant effect of [±Errors] on the three readability measures in the second experiment *Errors II*.Texts with [8% errors−Explicit] lead to shorter *Reading Times* than texts with [8% errors+Explicit] ($p = 0.027$), [50% errors short+Explicit] ($p = 0.010$), and [50% errors long+Explicit] ($p < 0.001$). Likewise, texts with [8% errors−Explicit] lead to shorter *Fixation Duration* than texts with condition [50% errors short+Explicit] ($p = 0.036$), and [50% errors long+Explicit] ($p = 0.007$). On the other hand, texts with [50% errors long+Explicit] condition lead to longer *Reading Times* than [50% errors short+Explicit] ($p < 0.001$), [8% errors+Explicit] ($p < 0.001$), and [8% errors−Explicit] ($p < 0.001$); Consistently, texts with [50% errors long+Explicit] lead to more fixations than text with conditions [8% errors−Explicit] ($p < 0.001$), [8% errors+Explicit] ($p < 0.001$), and [50% errors short+Explicit] ($p < 0.001$).

4.2 Comprehension

Table 3 shows the main statistical measures for the *Comprehension Score* for each of the conditions of the experiments *Errors I* and *II*. The results of the statistical tests show that:

- **Between Groups:** We found no effects of [±Errors] on the *Comprehension Score* between groups in the first experiment ($p = 0.54$), neither in the second experiment ($p = 0.170$).
- **Group D:** In the first experiment *Errors I* we could not find any effects of [±Errors] on *Comprehension Score* in Group D. However, in the second

Conditions	Group D		Group N		Ratio
	$\tilde{x}$	$\bar{x} \pm s$	$\tilde{x}$	$\bar{x} \pm s$	%
Errors I	*Comprehension Score*				
[No errors]	100	75.00 ± 42.96	100	**88.64 ± 30.60**	84.61
[16% errors−Explicit]	100	**77.27 ± 42.89**	100	68.18 ± 47.67	113.33
[16% errors+Explicit]	100	**72.73 ± 45.58**	100	59.09 ± 50.32	123.08
Errors II	*Comprehension Score*				
[8% errors−Explicit]	50	65.69 ± 35.34	100	**70.65 ± 37.00**	93.35
[8% errors+Explicit]	50	**61.76 ± 30.96**	50	49.20 ± 35.68	125.86
[50% errors short+Ex.]	100	87.25 ± 24.17	100	**96.08 ± 13.22**	90.61
[50% errors long+Ex.]	50	52.94 ± 35.23	50	**65.67 ± 35.02**	79.42

Table 3 Median, mean and standard deviation of the *Comprehension Score* for the *Errors* experiments. The comprehension score is normalized such that 100% is the value for people without dyslexia and no errors in the text.

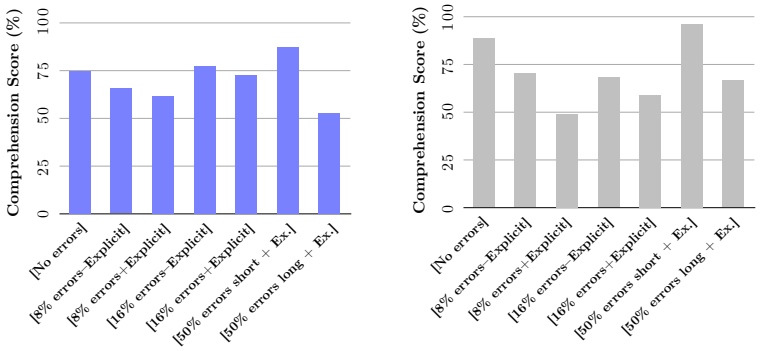

Fig. 5 *Comprehension Score* means for *Errors I* and *Errors II* experiments for Group D (left) and Group N (right).

experiment there was a significant effect of [±Errors] on the *Comprehension Score*. Texts with [50% errors short+Explicit] presented a significantly higher *Comprehension Score* than texts with [8% errors−Explicit] ($p = 0.004$), [8% errors+Explicit] ($p < 0.001$), and [50% errors long+Explicit] ($p < 0.001$).

- **Group N:** For Group N, there was a significant effect of [±Errors] on the *Comprehension Score* of both experiments. In *Errors I*, texts with [16% errors+Explicit] presented a significant lower *Comprehension Score* than text with [No errors] ($p = 0.080$).
  Likewise, *Errors II*, texts with [8% errors−Explicit] presented a significantly lower *Comprehension Score* than texts with [50% errors short+Explicit] ($p = 0.002$). Texts with [8% errors+Explicit] presented a significantly lower *Comprehension Score* than texts with [50% errors short+Explicit] ($p < 0.001$) and [50% errors long+Explicit] ($p = 0.034$). Texts with [50% errors short+Explicit] presented significantly higher *Comprehension Score* than texts with [50% errors long+Explicit] ($p = 0.010$).

	Group D	**Group N**	**Ratio**
	$\bar{x} \pm s$	$\bar{x} \pm s$	%
Errors I	*Reading Efficiency*		
[No errors]	4.30±4.27	**8.96**±7.40	48.00
[16% errors−Explicit]	3.71±3.64	**4.56**±3.74	81.36
[16% errors+Explicit]	3.06±2.82	**3.65**±3.89	83.90
Errors II	*Reading Efficiency*		
[8% errors−Explicit]	2.02±1.61	**2.66**±1.86	75.84
[8% errors+Explicit]	1.32±0.76	**1.72**±1.54	76.85
[50% errors short+Ex.]	1.85±1.10	**3.20**±1.30	57.81
[50% errors long+Ex.]	0.94±0.86	**1.34**±0.97	70.50

Table 4 Mean and standard deviation of the *Reading Efficiency* for the *Errors* experiments.

Errors II	**Group D**		**Group N**	
	$\tilde{x}$	$\bar{x} \pm s$	$\tilde{x}$	$\bar{x} \pm s$
	Readability Rating			
[8% errors−Explicit]	4	3.78 ± 0.78	4	**3.91 ± 1.00**
[8% errors+Explicit]	3	3.06 ± 1.07	3	**3.39 ± 1.16**
[50% errors short+Ex.]	4	**4.10 ± 0.81**	4	3.83 ± 0.99
[50% errors long+Ex.]	4	**3.37 ± 1.15**	4	3.25 ± 1.17
	Comprehensibility Rating			
[8% errors−Explicit]	4	4.02 ± 0.73	4	**4.39 ± 0.68**
[8% errors+Explicit]	4	3.06 ± 1.08	3	**3.59 ± 0.86**
[50% errors short+Ex.]	4	**4.20 ± 0.80**	4	**4.20 ± 0.81**
[50% errors long+Ex.]	4	3.39 ± 1.22	3.5	**3.57 ± 1.04**

Table 5 Median, mean and standard deviation of the *Readability Rating* and the *Comprehensibility Rating* for *Errors II*.

4.3 Reading Efficiency

For each participant, we also calculated his/her *Reading Efficiency*, defined as the *Comprehension Score* divided by the *Reading Time*. In Table 4 we show the means of the *Reading Efficiency* per condition and group. Last column represents the ratio of the *Reading Efficiency* of Group D divided by the one of Group N. This ratio allows us to compare both experiments, because even if the complexity of the texts read might differ, the relative performance among groups should not.

There was a significant effect of the *Reading Efficiency* between groups. Participants without dyslexia had a significantly higher *Reading Efficiency* ($\bar{x} = 3.15$, $s = 3.55$) than the participants with dyslexia ($\bar{x} = 2.08$, $s = 2.28$, $p < 0.001$).

4.4 Subjective Readability and Comprehensibility

There was a significant effect of [±Errors] on *Readability Rating* ($\chi^2(3) = 47.93$, $p < 0.001$), and on the *Comprehensibility Rating* ($\chi^2(3) = 77.15$, $p <$

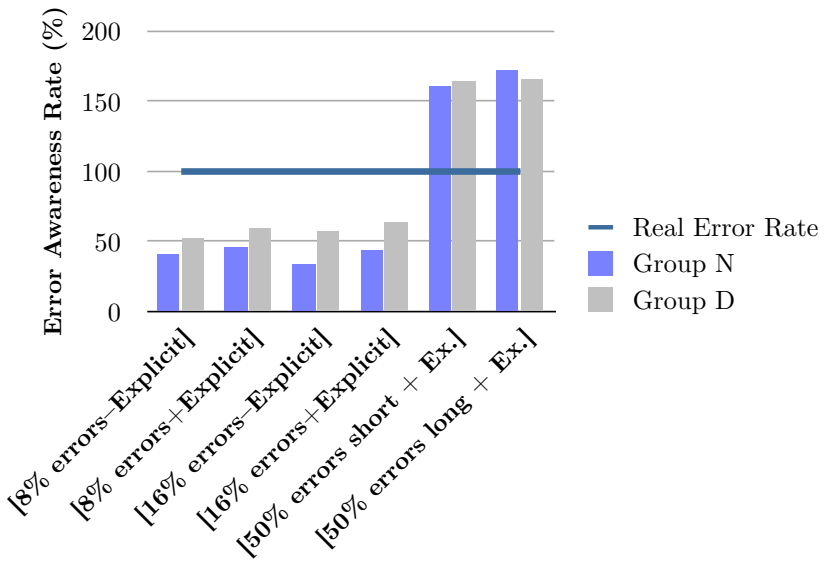

Fig. 6 Relative percentages of the *Error Awareness Rate* per condition and group in comparison with the real error rate.

0.001). Table 5 shows the main statistical measures obtained. The results of the post-hoc tests show that:

- **Between Groups:** We found no effects of [±Errors] on the *Readability Rating* between the groups ($p = 0.830$). However, there was a significant effect of [±Errors] on the *Comprehensibility Rating*. Participants without dyslexia found texts significantly more comprehensible ($\bar{x} = 3.94$, $s = 0.93$) than the participants with dyslexia ($\bar{x} = 3.67$, $s = 1.08$, $p = 0.011$).
- **Group D:** There were significant effects of [±Errors] on the *Readability Rating* and the *Comprehensibility Rating*. Texts with [8% errors+Explicit] were found to be less readable than texts with [8% errors−Explicit] ($p = 0.003$), and texts with [50% errors short+Explicit] ($p < 0.001$). Texts with [50% errors short+Explicit] were found to be more readable than texts with [50% errors long+Explicit] ($p = 0.005$).
  Likewise, texts with [8% errors+Explicit] were found to be less comprehensible than texts with [8% errors−Explicit] ($p < 0.001$), and text with [50% errors short+Explicit] ($p < 0.001$). Texts with [50% errors long+Explicit] were found to be less comprehensible than texts with [8% errors−Explicit] ($p = 0.047$), and [50% errors short+Explicit] ($p = 0.003$).
- **Group N:** For Group N, there was a significant effect of [±Errors] on the *Readability Rating* and the *Comprehensibility Rating*. Texts with [50% errors long+Explicit] were found less readable than text with [8% errors−Explicit] ($p = 0.025$). Similarly, texts with [8% errors+Explicit] were found to be less comprehensible than texts with [8% errors−Explicit] ($p < 0.001$), and [50% errors short+Explicit] ($p = 0.001$). Texts with [50%

errors long+Explicit] were found to be less comprehensible than texts with [8% errors−Explicit] ($p < 0.001$), and [50% errors short+Explicit] ($p = 0.004$).

4.5 Error Awareness

In Figure 6 we show the relative percentages of the error awareness rate for each group and condition in comparison with the real error rate. Even if we can observe in Figure 6 that Group N gets closer in their answers to the real error rate in the texts, we could not find any effect on the *Error Awareness Rate*: ($\chi^2(4) = 21.65,\ p < 0.001$) in *Errors I* and ($\chi^2(3) = 17.67,\ p = 0.001$) in *Errors II*. Hence the distribution of the *Error Awareness Rate* is not different among groups. Also, the Spearman's correlation coefficients between Group D and Group N on *Error Awareness Rate* are significant: $\rho = 0.799$ ($p < 0.001$) for *Errors I* and is $\rho = 0.470$ ($p < 0.001$) for *Errors II*.

5 Conclusions

The main conclusion is that in most of the cases:

- The presence of errors in the text does not impact the reading performance (readability and comprehension) of people with dyslexia, while it does for people without dyslexia.

This fact is shown in our experiments in several ways. First, regarding *Readability*, the reading time when the text has errors increases less in people with dyslexia. In fact, there are significant differences in the first experiment for people without dyslexia while for people with dyslexia there is not. On the other hand, in both experiments there were significant differences on the number of fixations for the people with dyslexia, implying that they do see the errors but process them in a different way.

Moreover, the *Comprehension* score does not seem to be affected by errors for people with dyslexia. In fact, in three of the comprehension tests of texts with errors people with dyslexia had better average comprehension performance, a fact that it is not found in literature when using correct text. The texts of the experiments were designed to have an equivalent complexity, however we computed the relative percentage comprehension ratio to make the experiments more comparable. This ratio is plotted in Figure 7 where the improvements for the people with dyslexia are clearly seen. That is, the presence of errors in a text is the first observed condition in which people with dyslexia achieve higher comprehension scores when reading than people without dyslexia. On the other hand, the letter transposition effect produces a similar reading and comprehension performance for both, people with and without dyslexia. That is, both groups read texts similarly with jumbled letters.

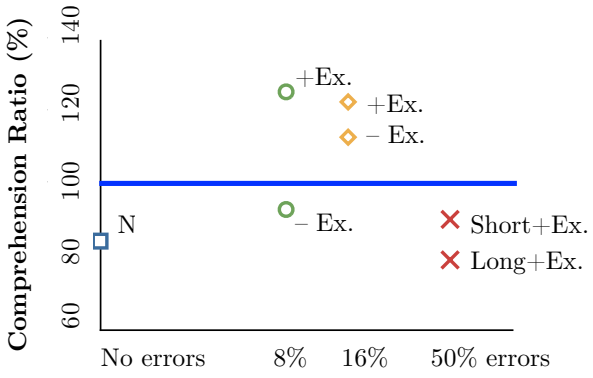

Fig. 7 Relative *Comprehension Score* percentage per condition for people with dyslexia. The comprehension score of people without dyslexia represents the line with value 100%.

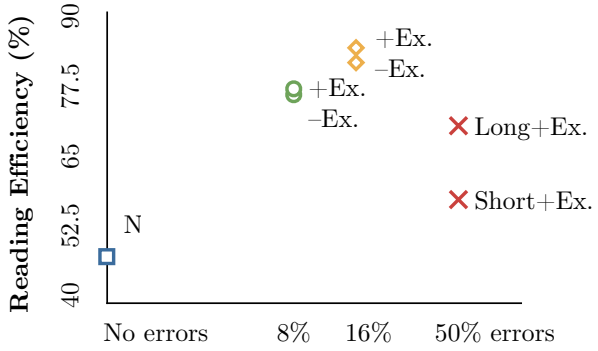

Fig. 8 *Reading Efficiency* of people with dyslexia percentage ratio per condition.

Finally, to put everything together we computed *Reading Efficiency*, that is, the comprehension score divided by the reading time. Here again we compute the percentage ratio that is shown in Figure 8. Clearly the reading efficiency with respect to the people without dyslexia increases reaching almost 90% for the case of 16% errors. Also, the reading efficiency is not affected by the knowledge of knowing that there are errors (both cases are almost in the same place for 8% and 16% errors), a reaffirmation of our main finding. Moreover, for the case of letter transposition the people with dyslexia improves even more for the case of long words, when normally long words take longer to process for people with dyslexia [31].

These results have theoretical and practical implications. First, they suggest that people with dyslexia process errors in a different way than people without dyslexia. Previous experiments on correct word recognition have shown how people with dyslexia present more difficulties than the rest. Until now, experiments on word recognition with people with dyslexia have not used written dyslexic errors or jumbled words. Further research involving this

condition would give insight on other parameters involved in dyslexic word recognition. Second, these findings can have an impact in how students with dyslexia are evaluated in the educational system. Written orthographic errors should be less penalized for people with dyslexia since there is evidence that people with dyslexia cannot (consciously) see the errors, having a lower error awareness rate that do not affect their comprehension.

Acknowledgements We thank Mari-Carmen Marcos for her assistance with the eye tracking equipment. We also thank all the participants that took part of the study.

References

1. Acha, J., Perea, M.: The effect of neighborhood frequency in reading: Evidence with transposed-letter neighbors. Cognition **108**(1), 290–300 (2008)
2. Bradford, J.: Designing web pages for dyslexic readers (2011). http://www.dyslexia-parent.com/mag35.html
3. British Dyslexia Association: Dyslexia style guide (2012). http://www.bdadyslexia.org.uk/
4. Bruck, M.: Word-recognition skills of adults with childhood diagnoses of dyslexia. Developmental Psychology **26**(3), 439 (1990)
5. Chapman, C.: The most popular fonts used by designers (2011). http://www.webdesignerdepot.com/2011/08/the-most-popular-fonts-used-by-designers/
6. Drndarevic, B., Saggion, H.: Towards automatic lexical simplification in Spanish: an empirical study. In: Proc. NAACL HLT '12 Workshop PITR '12. Montreal, Canada (2012)
7. Duñabeitia, J.A., Perea, M., Carreiras, M.: There is no clam with coats in the calm coast: Delimiting the transposed-letter priming effect. The Quarterly Journal of Experimental Psychology **62**(10), 1930–1947 (2009)
8. Evett, L., Brown, D.: Text formats and web design for visually impaired and dyslexic readers-clear text for all. Interacting with Computers **17**, 453–472 (2005)
9. Gregor, P., Newell, A.F.: An empirical investigation of ways in which some of the problems encountered by some dyslexics may be alleviated using computer techniques. In: Proc. ASSETS'00, ASSETS 2000, pp. 85–91. ACM Press, New York, NY, USA (2000)
10. Hyönä, J., Olson, R.K.: Eye fixation patterns among dyslexic and normal readers: Effects of word length and word frequency. Journal of Experimental Psychology: Learning, Memory, and Cognition **21**(6), 1430 (1995)
11. Inhoff, A.W., Rayner, K.: Parafoveal word processing during eye fixations in reading: Effects of word frequency. Perception & Psychophysics **40**(6), 431–439 (1986)
12. Johnson, R.L., Perea, M., Rayner, K.: Transposed-letter effects in reading: Evidence from eye movements and parafoveal preview. Journal of Experimental Psychology: Human Perception and Performance **33**(1), 209 (2007)
13. Just, M.A., Carpenter, P.A.: A theory of reading: From eye fixations to comprehension. Psychological review **87**, 329–354 (1980)
14. Lockley, S.: Dyslexia and higher education: accessibility issues. The Higher Education Academy (2002)
15. Mori, K., Komatsu, R.: Reading words with jumbled characters in Japanese. Perceptual and motor skills **112**(3), 947–950 (2011)
16. Olson, R.K., Kliegl, R., Davidson, B.J.: Dyslexic and normal readers' eye movements. Journal of Experimental Psychology: Human Perception and Performance **9**(5), 816 (1983)
17. Pedley, M.: Designing for dyslexics: Part 3 (2006). http://accessites.org/site/2006/11/designing-for-dyslexics-part-3-of-3
18. Pirozzolo, F.J., Rayner, K.: The neural control of eye movements in acquired and developmental reading disorders. Studies in Neurolinguistics **4**, 1–27 (1978)

19. Rack, J.P., Snowling, M.J., Olson, R.K.: The nonword reading deficit in developmental dyslexia: A review. Reading Research Quarterly pp. 29–53 (1992)
20. Rainger, P.: A dyslexic perspective on e-content accessibility (2012). www.texthelp.com/media/39360/USDyslexiaAndWebAccess.PDF
21. Raney, G.E., Rayner, K.: Word frequency effects and eye movements during two readings of a text. Canadian Journal of Experimental Psychology **49**(2), 151 (1995)
22. Rawlinson, G.: Reibadailty. New Scientist **162**(2188), 55–55 (1999)
23. Rawlinson, G.E.: The significance of letter position in word recognition. Ph.D. thesis, University of Nottingham (1976)
24. Rayner, K.: Eye movements and the perceptual span in beginning and skilled readers. Journal of experimental child psychology **41**(2), 211–236 (1986)
25. Rayner, K.: Eye movements in reading and information processing: 20 years of research. Psychological Bulletin **124**, 372–422 (1998)
26. Rayner, K., Duffy, S.A.: Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. Memory & Cognition **14**(3), 191–201 (1986)
27. Rayner, K., Raney, G.E.: Eye movement control in reading and visual search: Effects of word frequency. Psychonomic Bulletin & Review **3**(2), 245–248 (1996)
28. Rayner, K., White, S.J., Johnson, R.L., Liversedge, S.P.: Raeding wrods with jubmled lettres there is a cost. Psychological Science **17**(3), 192–193 (2006)
29. Rello, L., Baeza-Yates, R.: Good fonts for dyslexia. In: Proc. ASSETS'13. ACM Press, Bellevue, Washington, USA (2013)
30. Rello, L., Baeza-Yates, R.: How to present more readable text for people with dyslexia. Universal Access in the Information Society pp. 1–21 (2015)
31. Rello, L., Baeza-Yates, R., Dempere, L., Saggion, H.: Frequent words improve readability and short words improve understandability for people with dyslexia. In: Proc. INTERACT '13. Cape Town, South Africa (2013)
32. Rello, L., Baeza-Yates, R., Llisterri, J.: Dyslist: An annotated resource of dyslexic errors. In: Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014), pp. 1289–1296. Reykjavik, Iceland (2014)
33. Rello, L., Pielot, M., Marcos, M.C., Carlini, R.: Size matters (spacing not): 18 points for a dyslexic-friendly Wikipedia. In: Proc. W4A '13. Rio de Janeiro, Brazil (2013)
34. Serrano, F., Defior, S.: Dyslexia speed problems in a transparent orthography. Annals of Dyslexia **58**(1), 81–95 (2008)
35. Tressoldi, P.E., Stella, G., Faggella, M.: The development of reading speed in Italians with dyslexia a longitudinal study. Journal of learning disabilities **34**(5), 414–417 (2001)
36. Velan, H., Frost, R.: Cambridge university versus Hebrew university: The impact of letter transposition on reading English and Hebrew. Psychonomic Bulletin & Review **14**(5), 913–918 (2007)
37. White, S.J., Johnson, R.L., Liversedge, S.P., Rayner, K.: Eye movements when reading transposed text: the importance of word-beginning letters. Journal of Experimental Psychology: Human Perception and Performance **34**(5), 1261 (2008)
38. Williams, S., Reiter, E., Osman, L.: Experiments with discourse-level choices and readability. In: Proc. ENLG '03). Budapest, Hungary (2003)
39. Wimmer, H.: The nonword reading deficit in developmental dyslexia: Evidence from children learning to read German. Journal of Experimental Child Psychology **61**(1), 80–90 (1996)

Appendix

Misspellings used in the experiments. **treatro (teatro, 'theatre') (×2), *culturaleres (culturales, 'cultural'), *cerciendo (crescendo, 'growing'), *económinca (económica, 'economic'), *tendecia (tendencia, 'tendency'), *sumarion (sumaron, 'sum up'), *sectiores (sectores, 'sectors'), *fuertentemnte (fuertemente, 'strongly'), *criris (crisis, 'crisis'), *indutria (industria, 'industry'), *desperden (desprenden, 'clear'), *audiovisulares (audiovisuales, 'audiovisual'), *expectadores (espectadores, 'spectators') (×2), *venbidas (vendidas, 'sold'), *eurores (euros, 'euros'), *projectos (proyectos, 'projects'), *major (mayor,*

*'mayor'), *lso (los, ''the), *grades (grandes, 'big'), *esplicó (explicó, 'explained'), *segretario (secretario, 'secretary'),* and *genenral (general, 'general').*

–