The Information Technology and Innovation Foundation

Accessible Voting Technology Initiative

Working Paper Series

Working Paper #013

A Web Based Voting Application Study of Fonts for Voters with Dyslexia

Linda Harley, Keith Kline, Chandler Price, Adrienne Jones, Shaun Mosley, Sarah Farmer and Brad Fain

Georgia Tech Research Institute

December 2013

The Information Technology and Innovation Foundation
1101 K Street NW, Suite 610
Washington, DC 20005
(202) 449-1351

This research was supported by the U.S. Election Assistance Commission (EAC) under grant number EAC110149B. Any opinions, findings, conclusions or recommendations expressed in this report are those of the authors and do not necessarily represent the views of EAC or ITIF.

Preface

This report is part of a series of working papers. Some of the content of this report are identical to those of the other reports in the series, because the research studies had similar justifications and methods.

Executive Summary

The Help America Vote Act (HAVA) legislation was passed by Congress in 2002 in response to the controversy surrounding the 2000 U.S. presidential election. As a result many states changed their procedures and equipment to be more accessible and usable to voters who are hearing or visually impaired or who use a wheel-chair. However, there are many hidden disabilities that are not as apparent that should also be considered, such as individuals with dyslexia. Dyslexia is a reading disability that occurs when an individual has significant difficulty with the speed and accuracy of decoding words. This can affect their comprehension of the text and also lead to errors in spelling. Approximately 5-10% of people have dyslexia, but numbers vary depending on the research studies (Siegel, 2006). Reading is a central task for sighted voters, and dyslexia likely negatively impacts voters' experiences. Much of the text on ballots, such as candidate names and contest names, are short strings of text, but ballots sometimes also contain long paragraphs of text, such as in proposed constitutional amendments. Voters with dyslexia might take longer to read the short strings of text and the longer paragraphs. They also might make more errors than voters without dyslexia.

The goal of this research effort was to enable private and independent voting by voters with dyslexia. This research study **investigated three different font types**: Helvetica, Lexia Readable, and Open Dyslexic. Helvetica is a widely used sans-serif typeface. Lexia Readable is an adaption of Comic Sans, where letter symmetry is avoided and spacing between letters, words, and lines are increased. In Open Dyslexic font the letter strokes are thicker and vary within each letter, with heavier weight towards the bottom and left or right sides, to make pairs of commonly confused letters more distinguishable.

Twelve participants were recruited, seven with dyslexia and five without dyslexia. The order in which fonts were presented to participants was counterbalanced. The Voting App produced an event log that logged the time at which various events occurred. Eye tracking data was captured to determine gaze times. After using each font, participants were asked a series of questions. At the end participants were asked to rank the overall preference of the fonts.

No significant difference among font types were found for the objective data, that included race times, gaze times, and percent correct. However, the subjective data showed consistent difference preferences for Helvetica over Open Dyslexic or Lexia Readable. Participants with dyslexia reported that Helvetica was better than Open Dyslexic in regards to letter sharpness, letter legibility, and overall ease of reading. The only characteristic for which Helvetica was rated poorly was line spacing. The results indicate that for dyslexics and non-dyslexics alike, Helvetica with increased line spacing would be easier to read than Open Dyslexic or Lexia Readable.

Table of Contents

Preface 2
Executive Summary 3
Introduction 6
Font Types 6
Hypotheses 7
Method 8
Participants 8
Ballots 8
Eye Tracking System 9
Design 9
Procedure 11
Results 12
Dyslexia Screening 12
Vision 12
Dyslexia Questionnaire 12
Dyslexia Testing 13
Verbal Tests 14
Categorization of Subjects 16
Objective Data 17
Race Time 17
Prompt Time 18
Response Time 19
Percent Correct 20
Eye Tracking Results 21
Subjective Data 23
Ballot Ratings 23
Post Ballot Questionnaire 27
Discussion 28
Acknowledgments 29
References 30

List of Figures

Figure 1. Ballots with different fonts: (a) Helvetica, (b) Lexia Readable, and (c) Open Dyslexic. 7
Figure 2. SmartEye eye tracking system. 9
Figure 3. Self-reported dyslexia symptoms. 12
Figure 4. Average scores on the verbal tests. Diamonds represent means, boxes represent the interquartile range (25th to 75th percentile), and lines represent 1.5 times the interquartile range. 15
Figure 5. Standardized verbal test scores. Each line represents one participant's data. 16
Figure 6. Participants' average times on races with multiple candidates. Diamonds represent means. ... 17
Figure 7. Prompt times. 18
Figure 8. Normalized response time. 19
Figure 9. Percent Correct. 20
Figure 10. Average ratings of letter sharpness. 23
Figure 11. Average ratings of letter legibility. 24
Figure 12. Average rating of reading ease. 25
Figure 13. Average ratings of line spacing. 26

List of Tables

Table 1. Subjective rankings of fonts. Cell values represent number of participants.. 27

Introduction

Dyslexia is a disorder of phonological processing that causes difficulty in reading and writing (Shaywitz, Mody, & Shaywitz, 2006), and there have been many attempts to create fonts to facilitate reading for people with dyslexia(e.g., Bates, 2004; Gonzalez 2003; Leeuw de, 2010). Generally the designers of these fonts have attempted to make letters more distinguishable, particularly those letters that are often swapped by dyslexic readers, such as 'p' and 'q'. Letter symmetry and similarity may contribute to the difficulty that reader's with dyslexia experience.

Reading is a central task for sighted voters, and dyslexia likely negatively impacts voters' experiences. Much of the text on ballots, such as candidate names and contest names, are short strings of text, but ballots sometimes also contain long paragraphs of text, such as in proposed constitutional amendments. Voters with dyslexia might take longer to read the short strings of text and the longer paragraphs. They also might make more errors than voters without dyslexia.

Research suggests that certain fonts can mitigate difficulties reading for users with dyslexia (Leeuw de, 2010; Rello & Baeza-Yates, 2013). The present study was designed to evaluate the readability of three different fonts by voters with and without dyslexia. Although voters with dyslexia have the option of using an audible ballot, some may wish not to use it because of perceived difficulty, slowness, and social stigma.

Readers without dyslexia might find fonts for dyslexia unfamiliar and therefore more difficult to read than common fonts, such as Helvetica, Times, or Calibri. Therefore, there might be an interaction between dyslexia status and font type.

We expected non-dyslexic users to prefer Helvetica, and we expected dyslexic users to prefer either of two fonts that were developed for dyslexic readers: Open Dyslexic (Gonzalez 2003) and Lexia Readable (Bates 2004). We also expected voting and reading times to vary in accordance with the aforementioned preferences. Specifically, we expected voting and reading times to be shorter for voters with dyslexia using Open Dyslexic and Lexia Readable fonts, whereas we expected voters without dyslexia to have shorter voting and reading times with Helvetica.

Font Types

Helvetica, Lexia Readable, and Open Dyslexic fonts were used. Helvetica is a traditional, widely used font (Figure 1.a). Lexia Readable (Bates, 2004) is an adaption of Comic Sans; it is intended to appear less childlike (Figure 1.b). In an attempt to assist readers with dyslexia, the font designers used simple handwritten forms of certain letters, and letter symmetry was avoided. Spacing was increased between letters, words, and lines. Open Dyslexic font was also designed to make pairs of commonly confused letters more distinguishable (Figure 1.c). Stroke thickness varies within each letter, with heavier weight towards the bottom and left or right sides. This was intended to make letters such as 'p' and 'q' more distinct. The heavier weighting toward the bottom of letters was also intended to make text lines more prominent (Gonzalez, 2003).

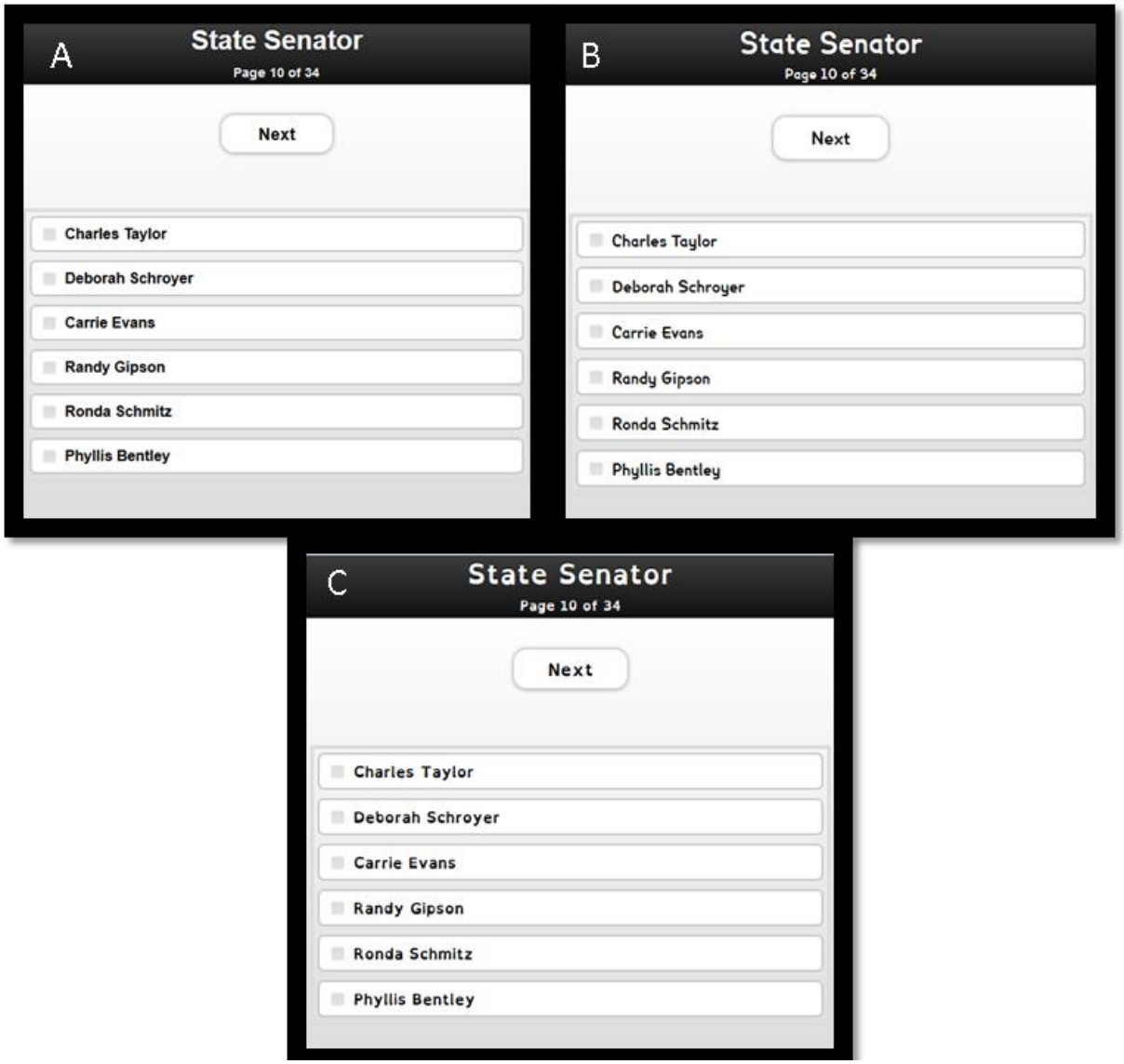

Figure 1. Ballots with different fonts: (a) Helvetica, (b) Lexia Readable, and (c) Open Dyslexic.

Hypotheses

The present study examined three different font types: Helvetica, Lexia Readable, and Open Dyslexic. These fonts were evaluated by users with and without dyslexia. The following hypotheses were tested:

H1. For objective measure of performance such as voting time and eye tracking data it was expected that non-dyslexic users would perform better on the Helvetica font, and dyslexic users would perform better on Open Dyslexic or Lexia Readable.

H2. For subjective user preferences it was expected that non-dyslexic users would prefer Helvetica font, while dyslexic users would prefer either Open Dyslexic or Lexia Readable.

Method

Participants

Twelve participants volunteered for this study. Participants without dyslexia were age and gender matched with participants with dyslexia as closely as possible. Self-reporting dyslexic participants included 5 males and 2 females (age = 39.4 ± 17.9 years). Non-dyslexic users included 4 males and 1 female (age = 34.4 ± 17.4). Participants were recruited from the Georgia Institute of Technology and surrounding Metro Atlanta area. Participants were compensated $50 for the 2 hour experiment. The Georgia Institute of Technology Institutional Review Board approved the study.

Ballots

Each participant was presented with three ballots on a computer. Each ballot was presented in a different font: Helvetica, Lexi Readable (LR), and Open Dyslexic (OD). Each ballot presented different candidates and amendments to minimize learning effects and familiarization with the ballot.

Each of the three ballots consisted of 13 items: 10 races, two retention questions (e.g., "Shall John Doe be kept as city comptroller?"), and one amendment. The ballots began with one practice race and ended with a review page. For the retention questions and amendment, participants were prompted to "vote yes or no on the next page." Participants were prompted to vote for a specific candidate immediately before each of the 10 races with competing candidates. In the list of candidates, the position of the target candidate was pseudo-randomized and balanced across the three ballots. Candidate name lengths were also matched across ballots. Amendment length and reading level were matched across ballots; their word lengths were 126, 149, and 154, with Flesch-Kincaid grade levels of 17.0, 15.9, and 16, respectively.

Three equivalent ballots were created for use with *each* of the three fonts. The order in which the fonts/ballots were presented was counterbalanced by Graeco-Latin square. This design also varied font-ballot pairings across participants. The three ballots were matched with regards to the locations of the correct (prompted) candidates in the races. For example, if ballot A had prompted candidates in list positions 3, 7, and 12, then ballots B and C did also. The locations of the matched races in the ballots were scrambled so that participants would not recognize a pattern as they proceeded through the experiment.

Over-votes and under-votes were not allowed by the Voting App (Harley et al., 2013a). Participants were not allowed to navigate back to a previous page so that they would be forced to recall and search for the name that had been primed. Similarly the "review" page button was disabled to force a linear progression through this ballot. These customizations of the Voting App were done for this study because the research question was interested in the readability of the ballot, and not the navigation or comprehension of voting mechanisms. Participants used a mouse to interact with the on-screen buttons.

Eye Tracking System

A SmartEye eye tracking system was used to track participants' direction of gaze as they voted. Three SmartEye cameras (Basler acA640-100gm cameras with 8mm lenses) with two IR flasher devices were used to sample eye position at 60Hz (Figure 2).

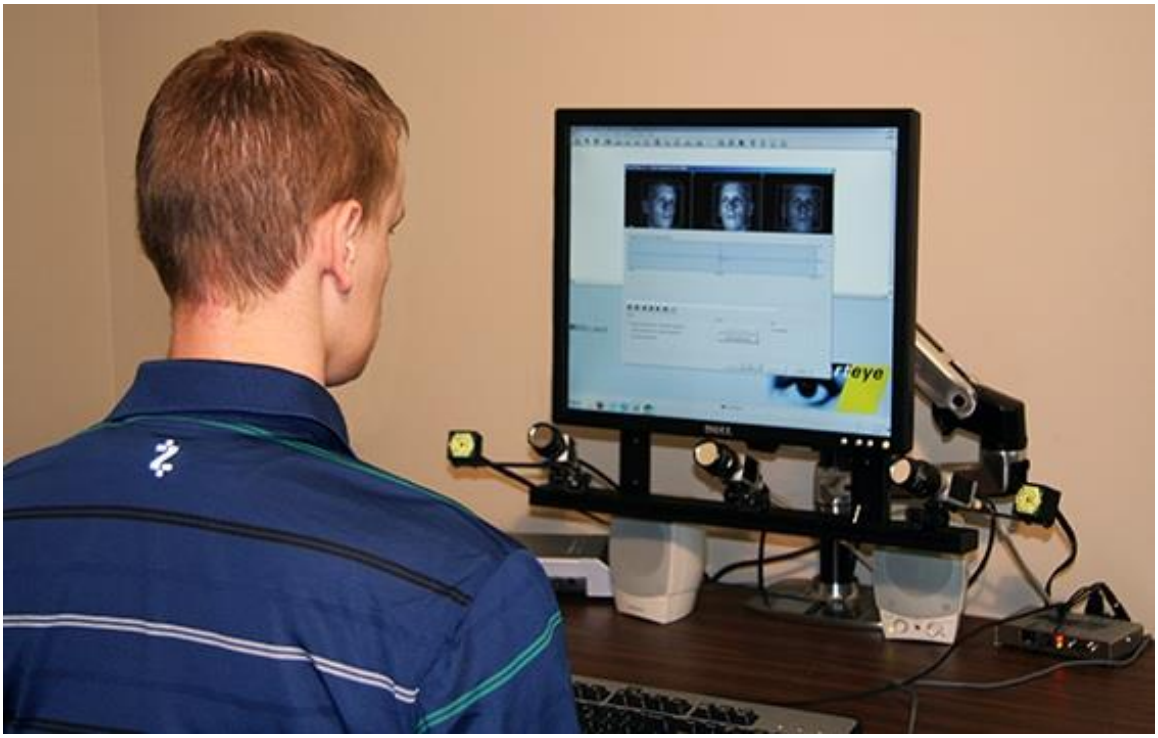

Figure 2. SmartEye eye tracking system.

Design

Dyslexia was the independent between participant grouping variable (participants with dyslexia; participants without dyslexia), and font type was a within participant independent variable (Helvetica; Lexi Readable (LR); Open Dyslexic (OD)) creating a 2 x 3 mixed factorial design.

Font type (Helvetica, LR, OD) was the independent variable for all analyses of user performance (correct selection of prompted candidates), user preference, and reading times. Reading times were estimated by two variables: (1) the amount of time the page was displayed and (2) the amount of time that a participant's gaze was directed towards a segment of text. Mixed ANOVAs were conducted on each dependent variable, with font type as a repeated measures variable and dyslexia status as a between subjects variable.

Immediately after using each ballot, participants who reported dyslexia completed ratings scale for the questions below. The questions were adapted from previous studies of colored overlays for dyslexic readers (Henderson, Tsogka, & Snowling, 2012). Response options included never, rarely, occasionally, frequently, and always.

- How often did the letters jumble?

- How often did you miss words?
- How often did you see words double?
- How often did the print seem to move?
- How often did the letters become fuzzy or blurry?
- How often did the space between the words form patterns like “rivers”?

Participants with dyslexia were also asked to rate the following qualities for each ballot, which were adapted from a study of fonts for dyslexic readers (Hillier, 2006):

- Spacing between letters (not enough, just right, too much)
- Spacing between words (not enough, just right, too much)
- Spacing between sentences (not enough, just right, too much)
- Spacing between lines (not enough, just right, too much)
- Spacing between paragraphs (not enough, just right, too much)
- Contrast (too low, just right, too high)
- Size of characters (too small, just right, too large)
- Legibility of the characters (not at all legible, slightly legible, barely legible, very legible)
- Ease of reading (very difficult, difficult, easy, very easy)

After using all ballots, participants ranked their readability.

Objective dependent variables were derived from event logs and eye tracking data. These provided estimates of readability by giving an indication of how long it took users to read the text.

Event logging variables included the following:

- *Race Time*: The time spent on the 10 multi-candidate races excluding the prompt pages (seconds). The initial instructions page, practice race, constitutional amendment, and ballot initiative were excluded from Race Time.
- *Prompt Time*: The time spent on the prompt pages that preceded each of the 10 multi-candidate races (seconds).
- *Response Time*: The time elapsed between race onset and candidate selection, divided by the candidate’s position in the list (seconds/candidate).
- *Percent Correct*: Percentage of candidate selections that matched the prompts.

Dependent variables for the eye tracking analysis included the following:

- *Instruction Gaze Time*: The amount of time spent looking at the instructions page at the beginning of each ballot (seconds).
- *Candidate Gaze Time*: The amount of time spent looking at the entire list of candidates (seconds).
- *Target Candidate Gaze Time*: The amount of time spent looking at the target candidate (seconds).

- *Visual Search Time*: The amount of time spent looking at all the options except the target candidate (seconds). This reflects the amount of time participants spent seeking the target candidate.
- *Retention Gaze Time*: The amount of time spent looking at the retention questions. There were two retention questions on each ballot.
- *Amendment Gaze Time*: The amount of time spent looking at the amendment texts. There were two amendments on each ballot.
- *Review Gaze Time*: The amount of time spent looking at the races and candidates on the review page.

Procedure

Participants completed the experiment in one-on-one sessions with an Experimenter, who was present throughout the session. Participants first signed an informed consent form and received a brief overview of the procedures from the experimenter. All participants were tested with the Rosenbaum near vision screening test to distinguish between poor vision and dyslexia. Participants who reported having dyslexia answered a short questionnaire regarding the impact of dyslexia on their daily activities. An experimenter administered the Revised Adult Dyslexia Checklist (Vinegrad, 1994) to all participants. An experimenter also administered a set of verbal assessment tests to all participants. This included a spoonerism test with 18 items and portions of the Wide Range Achievement Test 4 (Wilkinson & Robertson, 2006), including a word reading test, spelling test, and sentence comprehension test.

Next participants sat comfortably in front of a computer and eye tracking system. The experimenter calibrated the eye tracking system and then launched the Voting App. The participant completed the first ballot while the experimenter observed. Immediately afterward, the participant completed rating scales regarding readability and legibility of the ballot. The procedures were repeated for the next two ballots.

After all three ballots were completed, participants rank ordered the ballots by preference. A printed screenshot of each ballot was provided at this time. Participants also described the characteristics of each ballot that contributed to its readability and legibility. Lastly, participants were debriefed and compensated for their time.

Results

Dyslexia Screening

Vision

All participants had 20/70 vision or better, as determined by the Rosenbaum near vision screening test. Eight of the twelve participants reported that they wore glasses. Participants were allowed to wear their glasses during the study.

Dyslexia Questionnaire

Participants who reported having dyslexia answered a questionnaire about their experiences with dyslexia. The questions are listed below with summarized answers.

Participants were asked about the difficulties they experience when they read by selecting the symptoms from a list (Figure 3). The two most common responses listed by at least six participants were: (1) Difficulty understanding the different sounds that letters make, and (2) Reverse letters in words. Five participants reported: (1) can read to self but have difficulty reading aloud, (2) transpose syllables, (3) difficulty recognizing groups of letters as words.

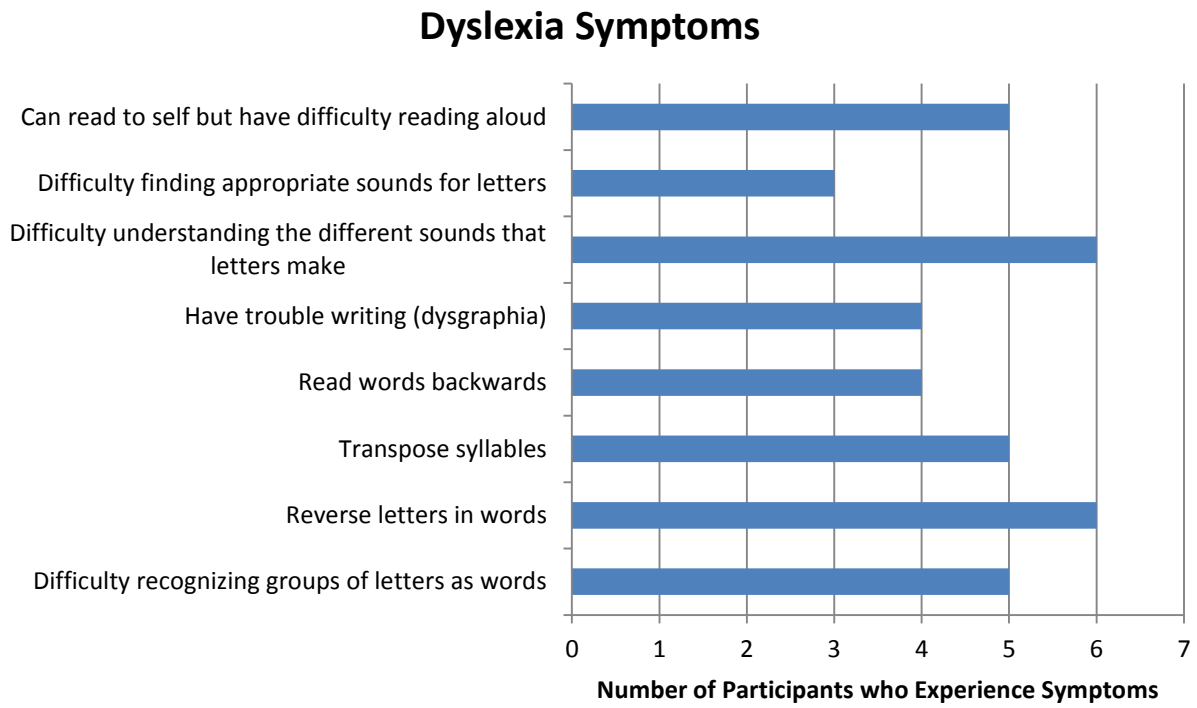

Figure 3. Self-reported dyslexia symptoms.

Please rate the severity of your dyslexia.

Three participants indicated that they had severe dyslexia. Two participants indicated that they had moderate dyslexia. One participant indicated slight dyslexia. One participant indicated dyslexia between slight and moderate.

Please rate the extent to which your dyslexia affects your daily life.

Four participants indicated that their dyslexia affected their daily lives an extreme amount. Two participants indicated that it had a moderate effect on daily life. One participant said daily life was slightly affected.

Which of the following reading tools do you currently use?

Two participants indicated using a reading pen. Two participants indicated using assistive software that reads aloud. Three participants use assistive software that changes the font, size, colors, spacing, or magnification of text. Two participants indicated using some other assistive device.

In the past, have you faced challenges while voting as a result of your dyslexia? For example, when using a paper ballot or an electronic ballot?

Three participants indicated that they have not faced challenges while voting as a result of their dyslexia. One participant had never voted, and three participants indicated they have faced challenges while voting as a result of their dyslexia. Two of these participants reported having severe dyslexia.

Are you aware that you can vote using an audio ballot at the polling place? If yes, have you ever used an audio ballot? If so, please describe how that affected your voting experience (easier, more difficult). If no, have you ever considered using an audio ballot? Why or why not?

Only one participant knew about the audio ballot method of voting and indicated that it made the voting experience easier. Five participants did not know about the audio ballot. Of those five, two indicated they might use the audio ballot, one indicated they absolutely would use it, and two would not use it.

Dyslexia Testing

All participants completed the Revised Adult Dyslexia Checklist (Vinegrad, 1994). Answering “yes” to any 9 of the 20 total questions is a strong indicator for dyslexia (this is the “full-scale”). Additionally, answering “yes” to any 7 of a select 12 questions is a second indicator of dyslexia (this is the “subscale” that consists of items that are more highly diagnostic of dyslexia).

For six of the seven self-reporting dyslexic participants, scores on both the full-scale and subscale Dyslexia Checklist indicated dyslexia. For the seventh participant, neither score indicated dyslexia. Four of the five participants without dyslexia had checklist scores that agreed with their self-reports. The fifth participant had partial agreement: the full-scale indicated dyslexia, but the subscale did not.

Verbal Tests

For each of the four verbal tests, t-tests were conducted to determine if there were significant differences in scores between the self-reporting dyslexic group and the non-dyslexic group.

Sentence Comprehension

The results of an independent samples t-test (equal variances assumed: Levene's Test $F = 1.78$, $\rho = 0.21$) revealed that there was not a significant difference on sentence comprehension scores between dyslexic participants and non-dyslexic participants, $t(10) = -1.792$, $p = .103$ (Figure 4).

Reading Test

The results of an independent samples t-test (equal variances assumed; Levene's Test $F = 0.23$, $\rho = 0.64$) indicated that there was a significant difference between the reading test scores, $t(10) = -2.38$, $\rho = 0.04$. Non-dyslexic participants scored higher than the dyslexic participants (Figure 4).

Spelling Test

The results of an independent samples t-test (equal variances assumed; Levene's Test $F = 1.61$, $\rho = 0.23$) indicated that there was a difference on spelling test scores, $t(10) = -2.339$, $\rho = 0.041$. Non-dyslexic participants scored higher on the spelling test than the dyslexic participants (Figure 4).

Spoonerisms

Levene's Test for Equality of Variances was significant, indicating unequal variances between the dyslexic and non-dyslexic scores on Spoonerisms, $F = 5.702$, $\rho = 0.038$. Therefore, the degrees of freedom in the following t-test were adjusted to correct for the inequality of variances. The results of an independent samples t-test (equal variances not assumed) indicated that there was a significant difference between the scores of the dyslexic participants and the non-dyslexic participants, $t(4.24) = -3.92$, $\rho = 0.015$. The dyslexic participants scored lower than the non-dyslexic participants (Figure 4).

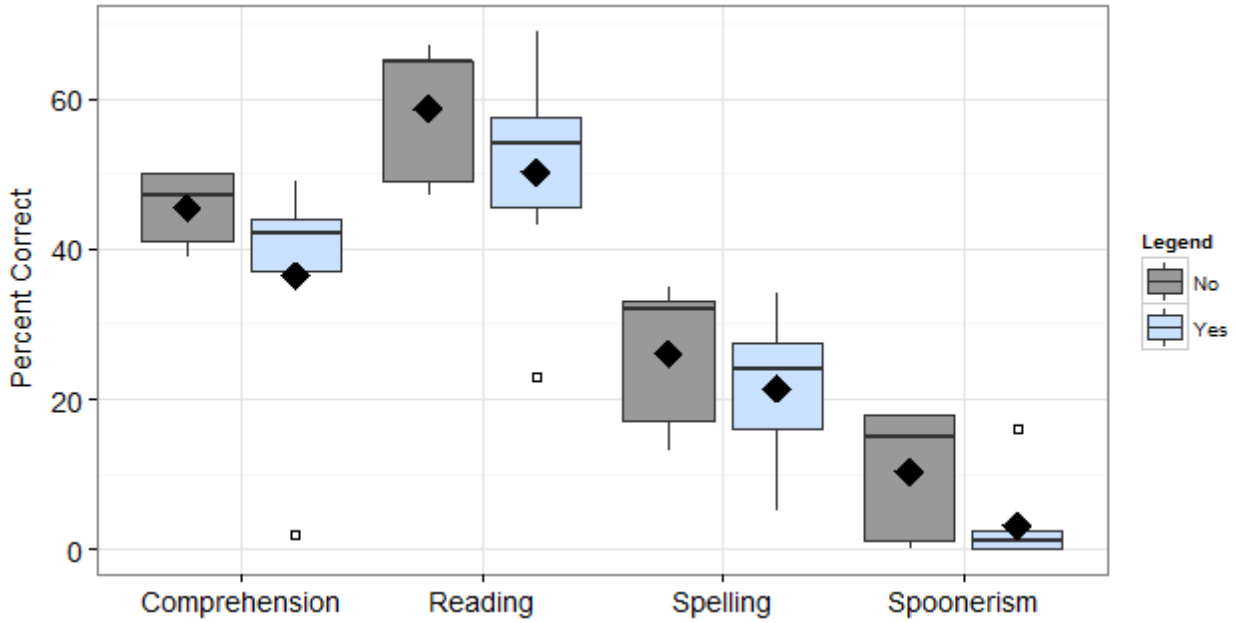

Figure 4. Average scores on the verbal tests. Diamonds represent means, boxes represent the interquartile range (25th to 75th percentile), and lines represent 1.5 times the interquartile range.

Categorization of Subjects

Although three of the four verbal tests showed statistically significant differences between groups, there was no clear dichotomy (or cut-off point) between the two groups on any of the four tests. Without a clear cutoff point, it was impossible to assign participants to the dyslexia and non-dyslexia groups on the basis of their scores. Therefore, subjects were divided into dyslexic and non-dyslexic groups solely on the basis of their self-report.

Note that self-reports agreed with Vinegrad Dyslexia Screening scores for all but one participant, whose score did not indicate dyslexia. This participant also scored high on the verbal tests

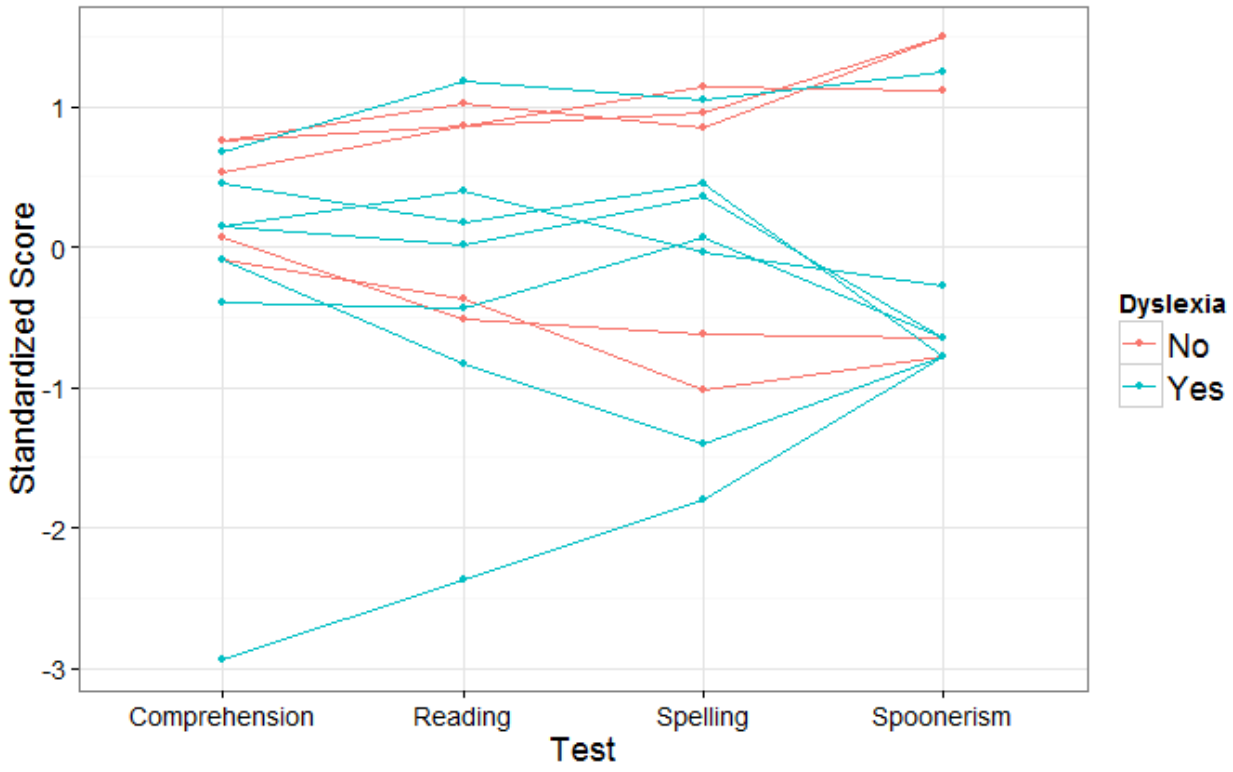

Figure 5. Standardized verbal test scores. Each line represents one participant's data.

Objective Data

Task and viewing durations were expected to be longer for the Helvetica ballot than the Lexi Readable and Open Dyslexic ballots for dyslexic participants (H1).

Race Time

Race time refers to the average amount of time participants spent on the races with multiple candidates. A split-plot ANOVA with dyslexia status as the between-subject variable and font type as the within-subject variable indicated that there was no significant effect of dyslexia status, $F(1, 10) = 0.21$, $\rho = 0.66$, $\eta_p^2 = 0.021$, power= 0.07, the font type, $F(2, 20) = 0.39$, $\rho = 0.68$, or the interaction of font type and dyslexia status, $F(2, 20) = 0.684$, $\rho = 0.52$ (Figure 6). However, the pattern of means was in the expected direction: Non-dyslexic users were faster with Helvetica ballot on average, and the dyslexic users were slightly faster with the Lexia Readable on average.

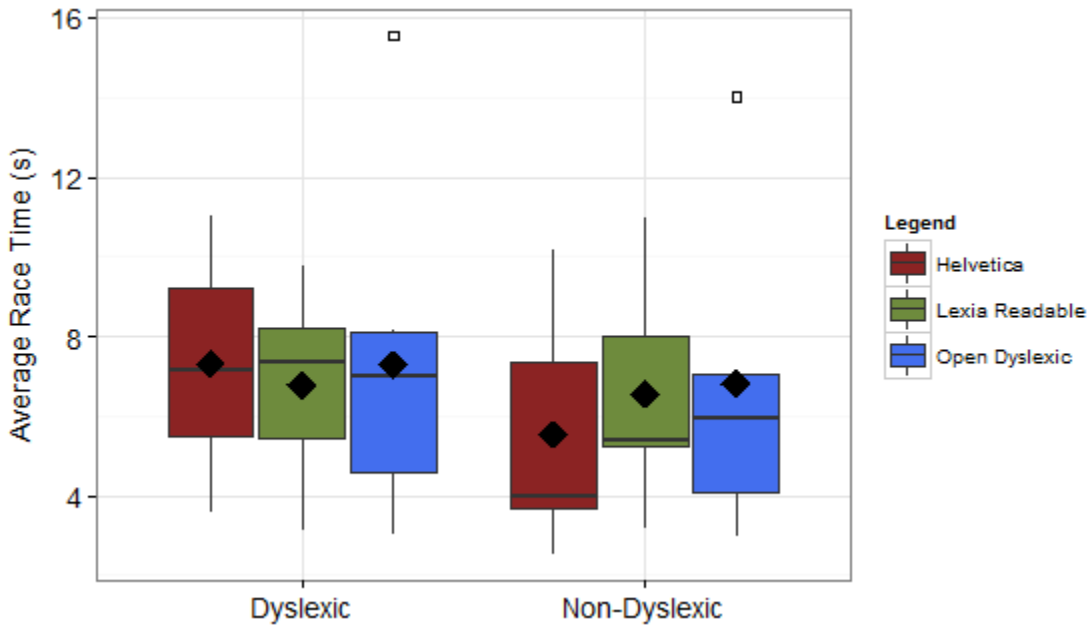

Figure 6. Participants' average times on races with multiple candidates. Diamonds represent means.

Prompt Time

Mean prompt time represents the average amount of time (in seconds) participants took to read the prompt (e.g., "vote for John Doe in the next race"). This was measured by the amount of time the prompt screen was displayed. A split-plot ANOVA with dyslexia status as the between-subjects variable and font type as the within-subjects variable indicated that there was no significant effect of dyslexia status, $F(1, 10) = 0.35$, $\rho = 0.57$, the font type, $F(2, 20) = 0.21$, $\rho = 0.81$, or the interaction of font type and dyslexia status, $F(2, 20) = 0.49$, $\rho = 0.62$ (Figure 7).

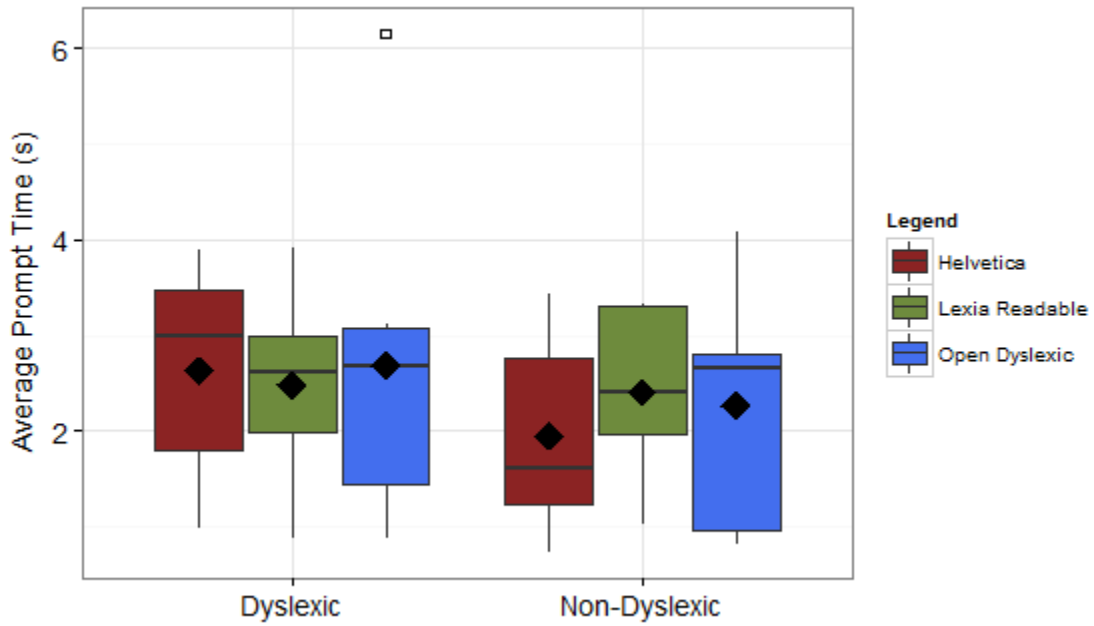

Figure 7. Prompt times.

Response Time

Mean response time indicates the average amount of time (in seconds) elapsed between the onset of the race and the candidate selection, excluding prompt time; this value was normalized by dividing it by the position of the candidate in the list. This controls for the number of candidates the participant must visually scan, and yields an average amount of time per candidate that elapsed between race onset and candidate selection, assuming the participant read the candidates on the screen sequentially from the top until finding the target candidate.

A split-plot ANOVA with dyslexia status as the between-subjects variable and font type as the within-subjects variable indicated that there was no significant effect of dyslexia status, $F(1, 10) = 0.52$, $\rho = 0.49$, the font type $F(2,20) = 0.2$, $\rho = 0.82$, or the interaction of font type and dyslexia status, $F(2, 20) = 1.36$, $\rho = 0.28$ (Figure 8). Again, although the differences in means were not statistically significant, the pattern of the means matched the prediction that non-dyslexic users would be faster with Helvetica, and the prediction that dyslexic users would be faster with Lexia Readable or Open Dyslexic.

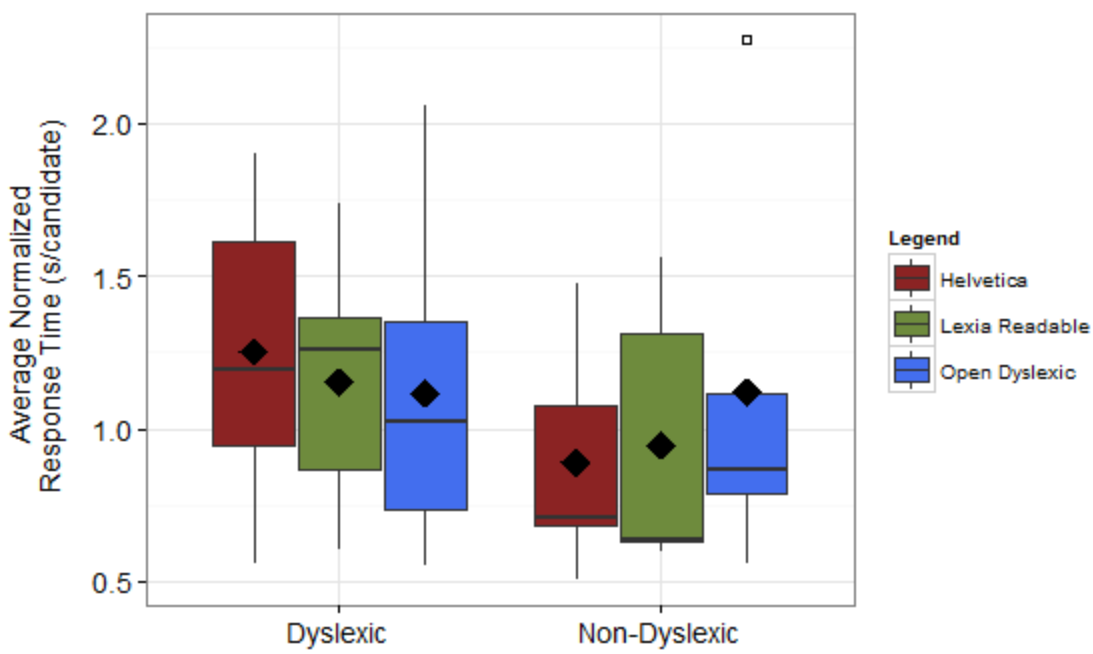

Figure 8. Normalized response time.

Percent Correct

Percent correct represents the percentage of correctly chosen candidates for each ballot (participants were told which candidate to vote for). A split-plot ANOVA with dyslexia status as the between-subjects variables and font type as the within-subjects variable indicated that there was no significant effect of dyslexia status, $F(1, 10) = 0.10$, $p = 0.72$, the font type, $F(2, 20) = 0.98$, $p = 0.37$, or the interaction of font type and dyslexia status, $F(2, 20) = 1.23$, $p = 0.32$ (Figure 9). On average, non-dyslexics made the most mistakes with the Lexia Readable ballot, while dyslexics made the most mistakes with the Open Dyslexic ballot.

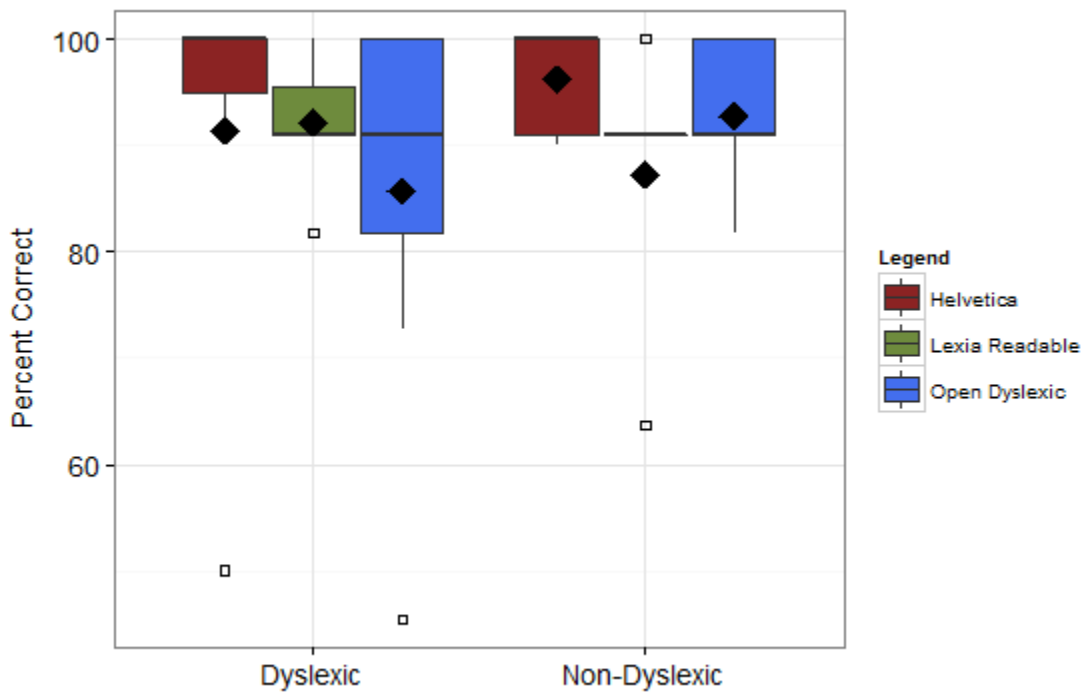

Figure 9. Percent Correct.

Eye Tracking Results

Dyslexic participants were expected to spend longer reading Helvetica than Open Dyslexic or Lexia Readable, but reading times did not differ significantly. The lack of statistical significance was likely due to insufficient power. Some effect sizes were medium-to-large, but power did not exceed 0.6 on any analyses of gaze times.

Instructions Gaze Time

Instruction gaze time referred to the amount of time participants spent looking at the instructions page at the beginning of each ballot. A split-plot ANOVA with dyslexia status as the between-subject variable and font types as the within-subjects variable indicated that there was no significant effect of dyslexia status, $F(1, 5) = 0.005$, $\rho = 0.945$, $\eta_p^2 = 0.001$, power= 0.05, the font type, $F(2, 10) = 0.315$, $\rho = 0.737$, $\eta_p^2 = 0.059$, power= 0.087, or the interaction of font type and dyslexia status, $F(2, 10) = 1.568$, $\rho = 0.256$, $\eta_p^2 = 0.239$, power= 0.257.

Candidates Gaze Time

Candidates gaze time refers to the amount of time participants spent looking at all candidate options on the ballot. A split-plot ANOVA with dyslexia status as the between-subjects variable and font type as the within-subjects variable indicated that there was no significant effect of dyslexia status, $F(1, 5) = 0.092$, $\rho = 0.774$, $\eta_p^2 = 0.018$, power= 0.057, the font type, $F(2, 10) = 0.251$, $\rho = 0.783$, $\eta_p^2 = 0.048$, power= 0.079, or the interaction of font type and dyslexia status, $F(2, 10) = 2.256$, $\rho = 0.155$, $\eta_p^2 = 0.311$, power= 0.354.

Target Candidate Gaze Time

Target candidate gaze time refers to the amount of time participants spent looking at the target candidate. A split-plot ANOVA with dyslexia status as the between-subjects variable and font type as the within-subjects variable indicated that there was no significant effect of dyslexia status, $F(1, 5) = 0.027$ $\rho = 0.876$, $\eta_p^2 = 0.005$, power= 0.052, the font type, $F(2, 10) = 0.436$, $\rho = 0.658$, $\eta_p^2 = 0.08$, power= 0.102, or the interaction of font type and dyslexia status, $F(2, 10) = 1.817$, $\rho = 0.212$, $\eta_p^2 = 0.267$, power= 0.293.

Visual Search Time

Visual search time refers to the amount of time spent looking at all the options except the target candidate. This reflects the amount of time participants spent seeking the target candidate. A split-plot ANOVA with dyslexia status as the between-subjects variable and font type as the within-subjects variable indicated that there was no significant effect of dyslexia status, $F(1, 5) = .153$, $p = .712$, $\eta_p^2 = 0.03$, power= 0.062, the font type, $F(2, 10) = 1.490$, $p = .271$, $\eta_p^2 = 0.23$, power= 0.246, or the interaction of font type and dyslexia status, $F(2, 10) = .741$, $p = .501$, $\eta_p^2 = 0.129$, power= 0.142.

Retention Gaze Time

Retention gaze time refers to the amount of time spent looking at the retention questions. A split-plot ANOVA with dyslexia status as the between-subjects variable and font type as the within-subjects variable indicated that there was no significant effect of dyslexia status, $F(1, 3) = 6.125$, $p = .09$, $\eta_p^2 = 0.671$, power= 0.402, the font type, $F(2, 6) = 3.478$, $p = .099$, $\eta_p^2 = 0.537$, power= 0.432, or the interaction of font type and dyslexia status, $F(2, 6) = 2.257$, $p = .186$, $\eta_p^2 = 0.429$, power= 0.298.

Amendment Gaze Time

Amendment gaze time refers to the amount of time spent looking at the amendment texts. A split-plot ANOVA with dyslexia status as the between-subjects variable and font type as the within-subjects variable indicated that there was no significant effect of dyslexia status, $F(1, 6) = .174$, $p = .691$, $\eta_p^2 = 0.74$, power= 0.065, the font type, $F(2, 12) = 3.040$, $p = .085$, $\eta_p^2 = 0.336$, power= 0.48, or the interaction of font type and dyslexia status, $F(2, 12) = 1.294$, $p = .310$, $\eta_p^2 = 0.177$, power= 0.228.

Review Gaze Time

Review gaze time refers to the amount of time spent looking at the races and candidates on the review page. A split-plot ANOVA with dyslexia status as the between-subjects variable and font type as the within-subjects variable indicated that there was no significant effect of dyslexia status, $F(1, 5) = .058$, $p = .820$, $\eta_p^2 = 0.011$, power= 0.055, the font type, $F(2, 10) = .712$, $p = .514$, $\eta_p^2 = 0.125$, power= 0.138, or the interaction of font type and dyslexia status, $F(2, 10) = 1.202$, $p = .341$, $\eta_p^2 = 0.194$, power= 0.206.

Subjective Data

Ballot Ratings

After each ballot was used, participants with dyslexia were asked to rate various aspects of the ballots (see Methods). A repeated measures ANOVA was performed on each of the 15 questions. Only the four questions showing significant differences among fonts are described below.

How often did the letters become fuzzy or blurry?

Five response options varied from "never" to "always". A repeated measures ANOVA revealed that the difference among the fonts was significant, $F(2, 22) = 4.331$, $p = .026$, $\eta_p^2 = 0.283$, power= 0.69 (Figure 10). Post-hoc tests using the Bonferroni correction revealed that the letters in the Open Dyslexic font became fuzzy or blurry more frequently than the letters in Helvetica, $\rho = 0.02$. There were no other significant differences.

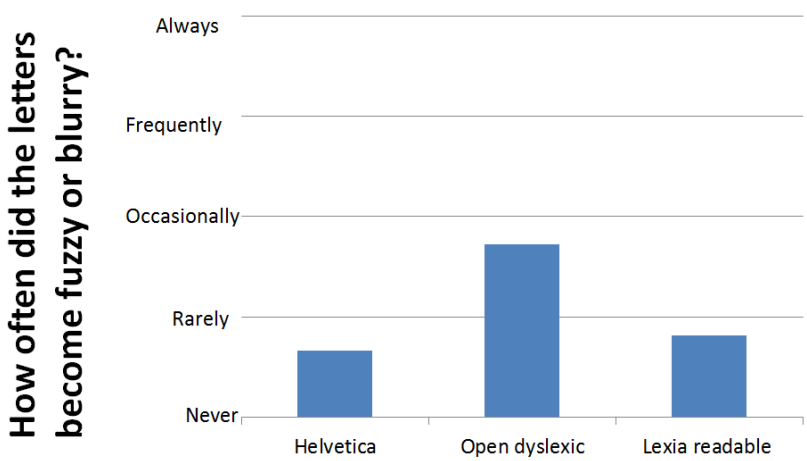

Figure 10. Average ratings of letter sharpness.

Legibility of characters

Four response options ranged from "not at all legible" to "very legible." A repeated measures ANOVA revealed that the difference among the fonts was significant, $F(2, 22) = 9.91$, $p = .001$, $\eta_p^2 = 0.474$, power= 0.968 (Figure 11). Post-hoc tests using the Bonferroni correction revealed that the legibility of the characters in the Helvetica ballot were more legible than those in the Open Dyslexic ballot, $p < .001$. There were no other significant differences.

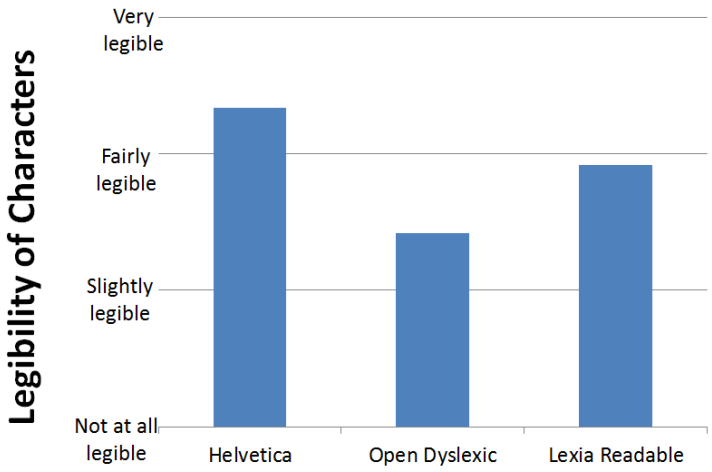

Figure 11. Average ratings of letter legibility.

Ease of reading

Four response options ranged from "very difficult to read" to "very easy to read." A repeated measures ANOVA revealed that the difference among the fonts was significant, $F(2, 22) = 5.04$, $p = .016$, $\eta_p^2 = 0.314$, power= 0.759 (Figure 12). Post-hoc tests using the Bonferroni correction revealed that the Open Dyslexic ballot was harder to read than either the Helvetic ballot, $p = .013$, or the Lexia readable ballot, $p = .017$.

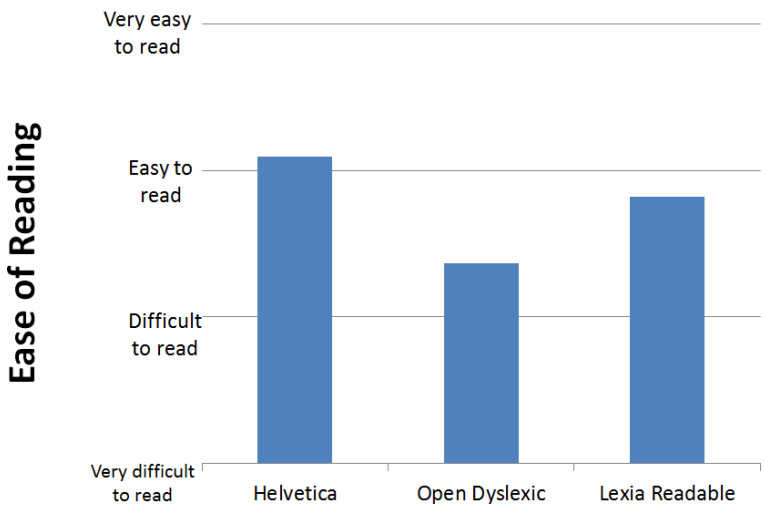

Figure 12. Average rating of reading ease.

Spacing of the lines was…

Response options included, "not enough space," "just the right amount of space," and "too much space." A repeated measures ANOVA revealed that the difference among the fonts was significant, $F(2, 22) = 5.10$, $p = .015$, $\eta_p^2 = 0.317$, power= 0.764 (Figure 13).Post-hoc tests using the Bonferroni correction revealed that the space between lines in the Lexia readable ballot was greater than that of the Helvetica ballot, $p = .017$. There were no other significant differences.

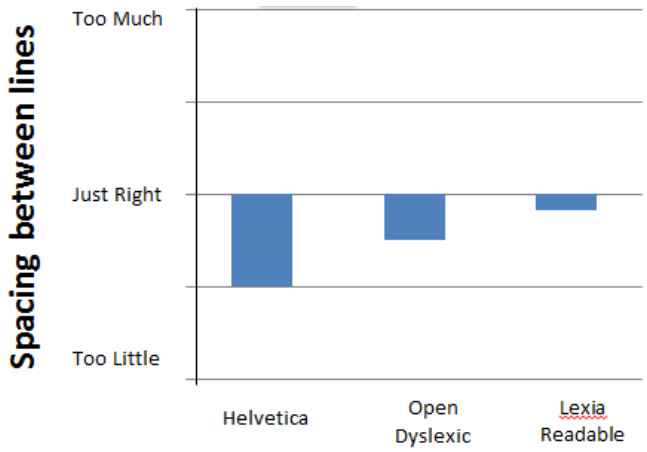

Figure 13. Average ratings of line spacing.

In summary, three of the four characteristics for which dyslexic participants gave significantly different ratings among fonts showed that Helvetica was better than Open Dyslexic. The fourth characteristic, line spacing, showed that Helvetica line spacing was significantly worse (too small) than Lexia Readable. Overall, Helvetica was the highest rated font, and it could easily be improved by increasing the line spacing.

Post Ballot Questionnaire

After all ballots were used, participants were asked to rank each font type on a scale from most legible to least legible. Friedman's test revealed there was a statistically significant difference in the ranking of the three font types, $X^2(2) = 12.167$, $p = .002$. The mean ranks for Helvetica, Lexia Readable, and Open Dyslexic were 2.75, 1.92, and 1.33, respectively (3 = most legible, 1 = least legible, see Table 1). Post-hoc analysis with Wilcoxon signed-rank tests revealed that the Helvetica font was ranked significantly higher than Open Dyslexic, $Z = -2.91$, $p = .004$, and Lexia Readable, $Z = -2.352$, $p = .019$. There was no significant difference in ranking between Open Dyslexic and Lexia Readable, $Z = -1.485$, $p = .138$.

Table 1. Subjective rankings of fonts. Cell values represent number of participants..

	Helvetica	Lexia Readable	Open Dyslexic
Most legible	9	2	1
Second most legible	3	7	2
Least legible	0	3	9
Total Score	33	23	16
Ratio Fav/Least Fave	infinity (9/0)	.67	.11

Participants were asked to describe the aspects of the fonts that contributed to their legibility. Answers are summarized below.

Helvetica: Most participants remarked that the Helvetica font was crisp and clear. Some also commented that it seemed more uniform and/or familiar than the other fonts.

Lexia readable: Several participants commented that this font had the most ideal spacing.

Open Dyslexic: Most of the comments for the Open Dyslexic font were commenting on its illegibility, e.g., that it was faded or blurry.

Participants were asked what they would recommend to improve the legibility of each of the three ballots (for example, the layout, color, and font). Answers are summarized below.

Helvetica: About half the participants had no recommendation for improvement. The other half noted that better spacing would improve this font.

Lexia readable: While participants enjoyed spacing between letters, some suggested increasing spacing between words because it was too similar to the distance between letters within each word.

Open Dyslexic: This font received a wide variety of comments for improving legibility, including several to completely change the font, change contrast, width, color, and/or spacing.

Discussion

The purpose of this research study was to determine which of three different fonts are preferred by users with and without dyslexia. The three fonts that were considered were; Helvetica, Lexia Readable and Open Dyslexic. It was hypothesized that users with dyslexia would prefer one of the fonts specially designed for them, while users without dyslexia would prefer Helvetica.

No significant differences among font types were found for objective variables, including display times, gaze times, and percent correct. However, with only one exception, the subjective variables showed consistent differences preferences for Helvetica over Open Dyslexic or Lexia Readable. Participants with dyslexia reported that Helvetica was better than Open Dyslexic in regards to letter sharpness, letter legibility, and overall ease of reading. The only characteristic for which Helvetica was rated poorly was line spacing. Dyslexic participants indicated that the lines were too close together – significantly more so than for Lexia Readable, which features expanded line spacing that is intended to improve readability for dyslexics.

It is somewhat ironic that the standard font, Helvetica, was preferred by dyslexics over two fonts that were developed with the express purpose of being easy to read for dyslexics. This is certainly not a condemnation of all fonts that have been developed for dyslexics, but it does present a cautionary message to developers: Font modifications that intuitively seem to improve legibility for dyslexics might fail to do so, and might decrease legibility instead.

The Open Dyslexic font featured thicker strokes in portions of letters that differentiate two mirrored letters (e.g., the descending lines in ‘p’ and ‘q’). Intuitively, it would seem that this might reduce the tendency for dyslexic readers to mistake the letters for each other. Several participants remarked that this effect was too extreme and made reading more difficult. This confirms Hillier’s (2006) finding that dyslexics prefer uniform strokes.

Several of the participants with dyslexia reported that they would use the accessible voting system, but were not aware of its functionality and capability. This is an example, of where a system designed for visually impaired individuals could also be utilized to assist voters with dyslexia. However, voter education and poll worker training does not typically account for this. This is one reason to incorporate screen reading capability with touchscreens, to aid users with dyslexia, and is a viable solution with today’s technology.

Overall, participants ranked Helvetica higher than Open Dyslexic and Lexia Readable. The results indicate that for dyslexics and non-dyslexics alike, Helvetica with increased line spacing would be easier to read than Open Dyslexic or Lexia Readable.

Acknowledgments

The authors would like to thank Dr. Cara Baily Fausset and Ms. Hannah Jahant for their contributions in reviewing and editing this manuscript.

References

Bates, K. (2004). Lexia Readable: K-Type.

Gonzalez, A. (2003). OpenDyslexic: Bitstream, Inc.

Henderson, L. M., Tsogka, N., & Snowling, M. J. (2012). Questioning the benefits that coloured overlays can have for reading in students with and without dyslexia. *Journal of Research in Special Educational Needs*.

Hillier, R. A. (2006). *A typeface for the adult dyslexic reader.* Anglia Ruskin University.

Leeuw de, R. (2010). *Special font for dyslexia?* , University of Twente.

Rello, L., & Baeza-Yates, R. (2013). *Good fonts for dyslexia.* Paper presented at the Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility.

Shaywitz, S. E., Mody, M., & Shaywitz, B. A. (2006). Neural mechanisms in dyslexia. [Article]. *Current Directions in Psychological Science, 15*(6), 278-281. doi: 10.1111/j.1467-8721.2006.00452.x

Vinegrad, M. (1994). A revised adult dyslexia checklist. *Educare, 48*(1), 21-23.

Wilkinson, G. S., & Robertson, G. (2006). Wide Range Achievement Test (WRAT4). *Psychological Assessment Resources, Lutz*.