Accelerating Adult Readers with Typeface: A Study of Individual Preferences and Effectiveness

Shaun Wallace
Brown University
Providence, RI 02912, USA
shaun_wallace@brown.edu

Ben D. Sawyer
University of Central Florida
Orlando, FL 32816, USA
sawyer@ucf.edu

Rick Treitman
Adobe Inc.
Newton, MA 02458, USA
treitman@adobe.com

Zoya Bylinskii
Adobe Inc.
Cambridge, MA 02142, USA
bylinski@adobe.com

Jeff Huang
Brown University
Providence, RI 02912, USA
chi@jeffhuang.com

CHI '20 Extended Abstracts, April 25–30, 2020, Honolulu, HI, USA.

ACM ISBN 978-1-4503-6819-3/20/04.
DOI: https://doi.org/10.1145/3334480.3382985

Abstract
Information overload is the challenge of the modern era and text the medium. Every adult reader would benefit from faster reading, provided they could retain comprehension. The present work explores the reading speed gains possible solely by manipulating typeface. We consider that optimal typeface might be a matter of an individual's preferred font, or that some fonts might be better for all users. Indeed, eight in ten of our participants believed their favorite font would be their best. Instead, our findings showed that the preferred font was seldom best, and one font did not fit all. Adult readers in our study read better with varying fonts. An average 117 word per minute difference between worst and best typeface, or around 10 additional pages an hour, means font choice is of real-world significance. Our discussion focuses on the challenges of rapidly identifying an individual's optimal font, and the exciting individuation technologies such an advance allows.

Author Keywords
reading, typography, preference, personalization

CCS Concepts
•**Human-centered computing** → **Human computer interaction (HCI);** *User studies;* Hypertext / hypermedia;

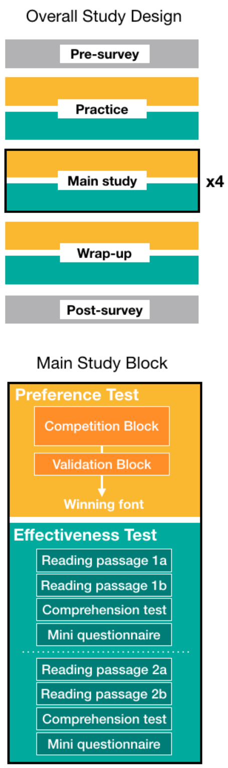

Figure 1: Our study alternated between preference and effectiveness tests. Each preference test included a double-elimination tournament of 4 fonts. In the effectiveness test, participants read passages in a fixed font which was either a random font or their most preferred font from the previous tournament.

Introduction and Related Work

The written word is the interface of choice for the information in our digital age. In work and pleasure alike, modern readers face pressure to comprehend ever-increasing amounts of information. In the face of information overload, new tools are needed. Prior work has identified one straightforward tool of optimization: typeface [1, 27]. In both at-a-glance and long form reading, font choice has been shown to mediate reading ability [4, 11, 15]. We posit that through optimal typeface, reading can be accelerated toward significant real-world improvement. If so, we expect a path toward systems which re-format information, optimizing the written word to enable fast, effective reading.

What, then, is an ideal font choice? While some fonts are indeed better at conveying information than others, subjectivity is commonplace. O'Donovan identified the difficulty modern users face selecting their preferred font [21]. If users could pick their preferred font, would it also be an effective font choice for them? Our work accepts this challenge to identify a user's most preferred font and compare it with their most effective font in terms of reading speed and comprehension. In the past, Boyarski et al. allowed users to make pairwise comparisons using two physical monitors, each showing a different font family [8]. This idea of pairwise comparisons to derive a definitive ranking for user preference is prevalent across the HCI community [17, 22, 29]. Recent work into font preference and effectiveness has mostly ignored this comparison method to derive a definitive ranking, instead relying on a Likert scale or ranking four or fewer fonts [3, 6, 7, 25, 28].

Starting with the hypothesis that people's font preferences can point to more effective fonts, in this work we design and validate a method to quickly determine a participant's preferred font using pairwise comparisons. Using our font preference toggle test, we can derive a user's definitive ranking from among 16 fonts. This method can be deployed in the wild, outside of lab settings, using crowd workers. We compare 16 modern fonts, many of which have never been studied before to evaluate user preferences and font effectiveness [3, 5, 6, 7, 8, 25].

This work offers a first step towards large-scale studies in users' naturalistic environments, to revisit the relationship between preference and readability. In this initial investigation, we introduce a preference toggle test and pit font preference against effectiveness, with a definitive ranking and a higher number of fonts than previous studies. Our evaluation metrics focus on the individual's experience, pointing towards findings that are personal in nature.

Procedure

Study design: To compare preference and effectiveness per font, we designed a study to alternate between (i) preference tests, where participants performed a toggle task comparing a set of four fonts and (ii) effectiveness tests to determine reading speed and comprehension, where participants read two sets of consecutive passages and answered comprehension questions in a single font (Fig. 1, bottom). Participants also completed a practice phase mimicking one main study block, and pre- and post-surveys. Study phases (i) and (ii) alternated per main study block and the wrap-up block (Fig. 1, top). In the wrap-up block, participants chose their most preferred font by comparing the four winning fonts from the main study blocks and then performed an effectiveness test with the same font. Participants saw all sixteen fonts across the preference tests. Due to study time limitations, participants only tested five total fonts for effectiveness. During the effectiveness tests, half of the time, participants read passages

PDF

Times

Arial

Calibri

Garamond

Newsprint

Poynter Gothic Text

Helvetica

Franklin Gothic

Utopia

Readability

Avant Garde

Avenir Next

Montserrat

Noto Sans

Web

Roboto

Open Sans

Lato

Oswald

Figure 2: The 16 fonts used during our study, chosen for their popularity and diverse use cases. Fonts are loaded from our server ensuring a consistent user experience. We used the open source EB Garamond.

in a preferred font (from their preference tests), and the other times in another randomly-assigned font.

Fonts: To account for fonts people encounter regularly, we selected 4 fonts from each of 4 sources (Fig. 2): (i) the most common fonts used for digital documents[1], (ii) popular fonts for print media [14], (iii) fonts recommended by readability experts [13, 26], and (iv) the most common fonts used on websites[2]. Our default font size of 16px is based on prior work [4, 12, 19] and is the default setting in modern browsers (e.g., Firefox, Chrome).

Reading passages: For the font preference tests, we used 10 passages (74–76 words) from a history textbook [18]. For the effectiveness tests, we obtained 12 text passages (113–175 words) from a reading specialist, on topics including nature, science, home improvement, etc. at an 8th-grade level, along with two comprehension questions per passage. During the study, we randomized, across participants, which passages were shown in which fonts.

Toggle task: For a user to converge on their preferred font quickly, we designed an interface to toggle between two fonts and have the user choose the preferred one, using the prompt: “What font is easier for you to read in?" (Fig. 3). Toggling between pairs of options at a time provides a simple and efficient method for assessment, motivated by other pairwise comparison tasks in the wild, such as eye exams and hearing aide adjustments [20].

Instead of an exhaustive comparison of all pairs of fonts, we use a double-elimination tournament, where a font is eliminated after a participant picks against it twice. The total number of pairwise comparisons using this approach is $(N-1) \times 2 + 1$, where N is the total number of fonts in this study. The pairing of fonts is randomized before each round of pairwise comparisons. In our study, participants also make repeat comparisons to validate the consistency of previous selections (Fig. 1, validation block). Our toggle task fundamentally differs from recent work that uses Likert-scale to measure preference [3, 24, 28].

Participants: We recruited 63 participants: 12 from university mailing lists, 15 from the UserTesting.com platform, and 36 from Amazon’s Mechanical Turk. Participants completed the study on the web using a device of their choice in their natural environments. We removed 3 university participants from the study because of unusually low comprehension scores or lack of comfort with English. Participants across all groups ranged in age from 18 to 55 years (average = 31). Overall, 51% of participants identified as female. It took 40 minutes on average to complete the study. Study compensation followed the pricing guidelines of each platform ($5–$20 per study).

Data preprocessing: Participants answered several voluntary pre-survey questions to ensure their data was not affected by any diagnosed disabilities (e.g., dyslexia), medical and neurological conditions (macular degeneration, diabetes, ADD, memory disorders, LPD, dyspraxia, etc.), and any other external factors (reading environment, caffeine, nicotine, etc.). For each participant who self-reported any of the above factors, we tested if their overall words-per-minute (WPM) or comprehension score fell outside the normal distribution of data using the interquartile range (IQR) method. Participants were also removed if their average dwell time per font during the preference test fell outside the normal distribution of data.

To establish a range of reading speed indicative of normal reading behavior, between skimming and reading for memorization, we extend Carver’s recommended range of

[1] Obtained from an Adobe corpus of 2302 PDF documents.

[2] https://fonts.google.com/analytics

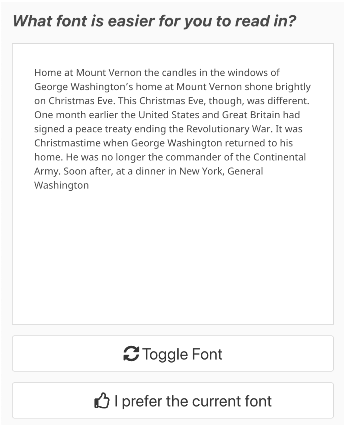

Figure 3: Participants toggle between pairs of fonts to decide which one is easier to read in. This interface has a fixed width of 420px regardless of the device. All text is rendered with a line spacing of 1.5. Custom JavaScript is used to ensure users cannot modify the size of the interface or text. This toggle test is done repeatedly within a double-elimination tournament over pairs of fonts to determine a participant's most preferred font. A participant toggles the font family used to display the sample text, then they stop on the font of the pair they prefer, and click to indicate their preference.

138–600 WPM, and remove any individual results outside 100–650 WPM [9, 10]. In this study, 9% of the participant data from the final 60 participants was removed based on the pre-processing methods described above.

Results and Discussion

Which are the highly-rated fonts? Noto Sans (chosen by 9 participants), Montserrat, and Garamond (each chosen by 8 participants) were selected most frequently as winners in the font preference tests (Table 1, 'Most Preferred'). The rest of the fonts were preferred by 1–5 participants each (except no one voted for Franklin Gothic). In other words, while there are clear winners, there is also diversity in font preferences across individuals.

Apart from the overall winners of the preference tests, we also considered the percent of pairwise match-ups each font won across all participants (Table 1, 'Win Rate'). For a fairer evaluation, 'Elo Rating' is a metric that takes into account the strength of the opponent in each pairwise match-up [16]. Win Rate and Elo Rating use many data points because each participant compared each font multiple times. According to these metrics, Noto Sans and Montserrat are the top fonts, but Garamond is in the bottom 5. How to reconcile this with the previous result? Garamond led to split opinions across participants - those who liked it, liked it a lot (it was their top font); others disliked it, voting it down in pairwise match-ups. Garamond has a high inter-participant disagreement score (Table 1, 'Disagreement'). For that matter, so does Montserrat, although it is still a top font, winning more pairwise match-ups against other fonts than it lost.

Overall, Noto Sans was a clear favorite. It was in the top 5 fonts for 46 participants (almost 80% of participants). With the highest Win Rate and average Elo Rating, it was the most consistently preferred font. Noto Sans has not been evaluated in previous studies [6, 8, 25].

Validation blocks during the font preference tests (Fig. 1) allowed us to measure participant response consistency. Individual preferences across unique font pairings were consistent 79% of the time. No participants were found inconsistent according to the IQR method.

Is familiarity with a font predictive of preference? We obtained ratings of participants' familiarity with all 16 fonts via a 3-point Likert scale administered during post-study surveys. We found no effects of font familiarity on preference, measured as Elo Rating ($r = -0.16, p < 0.017$, Bonferroni corrected). The most preferred font, Noto Sans, was also among the least familiar fonts to participants. These findings contradict prior work showing predictive relationships between font familiarity and preference [2].

What were the most effective fonts? We used two metrics to capture effectiveness: reading speed - measured in words-per-minute (WPM), and comprehension score - as the proportion of comprehension questions answered correctly after reading passages in a particular font. A font can not be deemed effective only if the reading speed increases. A high WPM but low comprehension score could be an indication of skimming: participants continuing to the next reading screen without internalizing the content. We observe high WPM paired with a low comprehension score for Garamond and Oswald. We assume a font is effective if both WPM and comprehension are high, which is the case for Noto Sans and Lato. Interestingly, the fonts that were most ineffective for reading, having both the lowest WPM and worst comprehension scores, were Roboto and Avant Garde. However, there were not enough data points for the pairwise differences in WPM between fonts to be statistically significant at the $p = 0.05$

Table 1 columns:
Most preferred: The total number of participants (out of 60) who selected the font as their absolute favorite.
Win Rate: The percent of pairwise match-ups each font won during the font preference test.
Elo Rating: To account for the strength of each font in a pairwise comparison during the double-elimination tournament, an Elo Rating [16] was computed per font per participant, and averaged across participants. The initial Elo Rating per font was 1500, and the system ran with $K = 64$.
Disagreement: The standard deviation across participants' Elo Ratings per font. The greater the number, the less consensus there was among overall preference per font.
Font Familiarity: 3-point Likert scale question from the post-survey (3 = extremely).
Times Read: Number of times a font was tested (read).
Words per Minute: WPM is averaged across passages and participants.
Comprehension: Average accuracy at answering four comprehension questions after reading in a particular font.

Font	Most Preferred	Win Rate	Average Elo Rating	Disagreement	Font Familiarity	Times Read	Words per Minute	Comprehension
Noto Sans	9	62%	1635	88	1.82	70	289	94.4%
Montserrat	8	56%	1598	124	1.79	102	285	91.7%
Garamond	8	43%	1421	124	1.88	36	339	83.3%
Roboto	5	56%	1583	78	1.84	83	254	85.0%
Lato	4	54%	1553	73	1.77	76	285	92.8%
Helvetica	4	53%	1525	103	2.30	62	315	86.6%
Arial	4	55%	1567	88	2.64	54	296	83.9%
Poynter Gothic	3	54%	1542	63	1.77	112	279	94.4%
Times	3	47%	1471	99	2.39	193	287	93.5%
Avenir Next	3	48%	1502	82	1.80	57	278	86.7%
Utopia	3	51%	1505	90	1.91	66	282	85.9%
Open Sans	2	57%	1598	71	2.00	73	300	85.5%
Oswald	2	22%	1236	110	1.80	19	315	85.0%
Calibri	1	46%	1512	80	2.27	65	311	88.2%
Avant Garde	1	40%	1422	142	1.86	47	271	85.4%
Franklin Gothic	0	34%	1329	78	1.82	48	312	86.5%

Table 1: Results show Noto Sans performed consistently well. It was the Most Preferred, with highest Win Rate and Elo Rating, and the most effective font for comprehension across all 16 fonts, while maintaining average WPM (*see metric descriptions in left-hand column*).

level, using two-tailed t-tests with Bonferonni correction. The WPM of the fastest 5 fonts (average = 321) taken together, *was* significantly higher than the average WPM of the slowest 5 fonts (average = 281, p < 0.003).

We conducted One-Way ANOVAs to assess the differences in reading speed and comprehension across the three user populations tested. Students from the university population read faster (avg. WPM = 320, SD: 137; $F(2, 1163) = 29, p < 0.01$) than crowd workers (User Testing avg. WPM = 303, SD: 124; MTurk avg. WPM = 278, SD: 116). However, there was not a significant difference in reading comprehension among the three populations.

What factors drive reading speed? The two fonts with the highest WPM were Garamond and Oswald. Garamond has the smallest x-height and Oswald has the second smallest width across all fonts in our study, potentially reducing the reading span. A piece of text split across fewer lines can be read faster because moving the eyes from the end of one line to the beginning of another slows down reading [23]. However, given that Noto Sans and Lato also have high WPM despite being pretty large fonts, font size can not be the only factor driving reading speed.

We used 'Mini questionnaires' (Fig. 1) to get participants to rate their familiarity with, and interest in, the text pas-

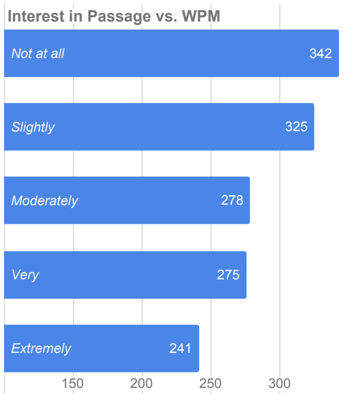

Figure 4: Participants slowed down when they read passages that they found interesting. Reading speed was measured in words per minute (WPM), while the interest in the passage was rated on a 5-point Likert scale after the reading and comprehension questions were completed.

sages read, using 5-point Likert scales. Contrary to our initial assumption, the more interesting a passage was rated, the slower it was read (Fig. 4). An exciting question for future investigation is whether a good reading font, by facilitating engagement with the text, can increase user's perceived interest in a reading passage.

Does familiarity with a font predict effectiveness? We found no effects of font familiarity on reading speed ($r = 0.042$) or comprehension ($r = 0.033$). In summary, familiarity with a font was predictive *neither* of preference *nor* effectiveness of that font.

Is preference predictive of effectiveness? We return to the question that motivated this whole study. To guide users to their most effective font, it would be easiest if their most preferred font was also their most effective. We measured whether participants reading in their preferred fonts were more effective, in reading speed and comprehension, compared to when reading in other fonts.

Participants read in their most preferred font at an average rate of 303 WPM. In contrast, they read in their fastest font at an average rate of 347 WPM (15% faster). Only 18% of participants read the fastest in their most preferred font; 23% read the slowest in their most preferred font (among the 5 fonts tested). Given that participants read in their slowest font at an average rate of 230 WPM, there is a 51% difference in individual reading speed between slowest vs. fastest fonts. Around 59% of participants scored the highest reading comprehensions scores after reading in their preferred font; 41% scored the lowest with their preferred font.

Participants do not know what is good for them. While the preference test succeeded in guiding people to their most preferred font (92% of participants agreed with their final font recommendation), it did not always guide them to their most effective font. These results run contrary to the belief among 80% of participants that their most preferred font would also be their most effective font to read in.

Because we found no consistency across participants in which fonts were most effective, this points to the future need for designing individuated reading experiences that account for individual differences. Moreover, we can not count on individual preferences for effective font selection.

Conclusion and Future Work

People read 51% faster in their fastest font compared to their slowest font. This translates to potential gains of 10 additional pages of reading per hour! Given these gains and our finding that different fonts are effective for different people, there is an exciting opportunity to build custom reading experiences and augment reading performance for adult readers.

Unfortunately, an individual's preference for a font does not predict their reading speed in that font. Discovering which behavioral patterns or individual differences (e.g., age, reading experience, education level, eye conditions, etc.) can help rapidly identify the most effective font for an individual is a promising future direction. For instance, some initial qualitative observations point to younger participants actually preferring and reading faster with smaller fonts. Additionally, whether a preferred font can create a more favorable experience and encourage someone to read more is a question for future investigation.

The potential impacts on individual reading efficacy highlighted here point to a future where machines help adult readers to reach their full reading potential. We invite the multidisciplinary communities that will perform this work to join us. Let us engineer better reading for everyone.

REFERENCES

[1] Aries Arditi. 2004. Adjustable typography: an approach to enhancing low vision text accessibility. *Ergonomics* 47, 5 (2004), 469–482.

[2] Justin Baer, Mark Kutner, John Sabatini, and Sheida White. 2009. Basic Reading Skills and the Literacy of America's Least Literate Adults: Results from the 2003 National Assessment of Adult Literacy (NAAL) Supplemental Studies. NCES 2009-481. *National Center for Education Statistics* (2009).

[3] Jayeeta Banerjee, Deepti Majumdar, Madhu Sudan Pal, and Dhurjati Majumdar. 2011. Readability, subjective preference and mental workload studies on young indian adults for selection of optimum font type and size during onscreen reading. *Al Ameen Journal of Medical Sciences* 4, 2 (2011), 131–143.

[4] Michael Bernard, Chia Hui Liao, and Melissa Mills. 2001. The effects of font type and size on the legibility and reading time of online text by older adults. In *CHI'01 extended abstracts on Human factors in computing systems*. ACM, 175–176.

[5] Michael Bernard, Bonnie Lida, Shannon Riley, Telia Hackler, and Karen Janzen. 2002. A comparison of popular online fonts: Which size and type is best. *Usability news* 4, 1 (2002), 2002.

[6] Michael L Bernard, Barbara S Chaparro, Melissa M Mills, and Charles G Halcomb. 2003. Comparing the effects of text size and format on the readibility of computer-displayed Times New Roman and Arial text. *International Journal of Human-Computer Studies* 59, 6 (2003), 823–835.

[7] Sanjiv K Bhatia, Ashok Samal, Nithin Rajan, and Marc T Kiviniemi. 2011. Effect of font size, italics, and colour count on web usability. *International journal of computational vision and robotics* 2, 2 (2011).

[8] Dan Boyarski, Christine Neuwirth, Jodi Forlizzi, and Susan Harkness Regli. 1998. A study of fonts designed for screen display. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM Press/Addison-Wesley Publishing Co., 87–94.

[9] Ronald P Carver. 1990. *Reading rate: A review of research and theory.* Academic Press.

[10] Ronald P Carver. 1992. Reading rate: Theory, research, and practical implications. *Journal of Reading* 36, 2 (1992), 84–95.

[11] Maneerut Chatrangsan and Helen Petrie. 2019. The effect of typeface and font size on reading text on a tablet computer for older and younger people. In *Proceedings of the 16th Web For All 2019 Personalization-Personalizing the Web*. ACM, 5.

[12] Jan Constantin. 2013. Typographic Design Patterns And Current Practices (2013 Edition) — Smashing Magazine. `https://www.smashingmagazine.com/2013/05/typographic-design-patterns-practices-case-study-2013/`. (May 2013). (Accessed on 09/19/2019).

[13] Kathy Crowley and Marjorie Jordan. 2019. Base Font Effect on Reading Performance - Readability Matters. `https://readabilitymatters.org/articles/font-effect`. (June 2019). (Accessed on 09/19/2019).

[14] Bill Davis. 2004. Fonts on the front page. A study of typefaces on the front pages of American's top newspapers. *Chicago: Ascender Corporation* (2004).

[15] Jonathan Dobres, Nadine Chahine, Bryan Reimer, David Gould, Bruce Mehler, and Joseph F Coughlin. 2016. Utilising psychophysical techniques to investigate the effects of age, typeface design, size and display polarity on glance legibility. *Ergonomics* 59, 10 (2016), 1377–1391.

[16] Arpad E Elo. 1978. *The rating of chessplayers, past and present*. Arco Pub.

[17] Shengbo Guo and Scott Sanner. 2010. Real-time multiattribute bayesian preference elicitation with pairwise comparison queries. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. 289–296.

[18] E. D. Hirsch Jr. (Ed.). 2017. *Early Presidents and Social Reformers*. Core Knowledge Foundation. https://www.coreknowledge.org/wp-content/uploads/2017/09/CKHG_G4_B3_EarlyPresSocRef_SR.pdf

[19] Aliaksei Miniukovich, Antonella De Angeli, Simone Sulpizio, and Paola Venuti. 2017. Design guidelines for web readability. In *Proceedings of the 2017 Conference on Designing Interactive Systems*. ACM, 285–296.

[20] Arlene C Neuman, Harry Levitt, Russell Mills, and Teresa Schwander. 1987. An evaluation of three adaptive hearing aid selection strategies. *The Journal of the Acoustical Society of America* 82, 6 (1987), 1967–1976.

[21] Peter O'Donovan, Jānis Lībeks, Aseem Agarwala, and Aaron Hertzmann. 2014. Exploratory font selection using crowdsourced attributes. *ACM Transactions on Graphics (TOG)* 33, 4 (2014), 92.

[22] Li Qian, Jinyang Gao, and HV Jagadish. 2015. Learning user preferences by adaptive pairwise comparison. *Proceedings of the VLDB Endowment* 8, 11 (2015), 1322–1333.

[23] Keith Rayner, Timothy J Slattery, and Nathalie N Bélanger. 2010. Eye movements, the perceptual span, and reading speed. *Psychonomic bulletin & review* 17, 6 (2010), 834–839.

[24] Luz Rello and Ricardo Baeza-Yates. 2016. The effect of font type on screen readability by people with dyslexia. *ACM Transactions on Accessible Computing (TACCESS)* 8, 4 (2016), 15.

[25] Luz Rello, Martin Pielot, and Mari-Carmen Marcos. 2016. Make it big!: The effect of font size and line spacing on online readability. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 3637–3648.

[26] Bonnie Shaver-Troup, Kathy Crowley, and Marjorie Jordan. 2017. Optimizing Reading Performance by Manipulating the Shape, Size, and Spacing of Text to Match the Individual's Visual Processing Capacity. *Collective Impact Project, revReading* (2017).

[27] Miles Albert Tinker. 1963. *Legibility of print.* Technical Report.

[28] Junxiang Wang, Jianwei Yin, Shuiguang Deng, Ying Li, Calton Pu, Yan Tang, and Zhiling Luo. 2018. Evaluating User Satisfaction with Typography Designs via Mining Touch Interaction Data in Mobile Reading. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 113.

[29] Jinfeng Yi, Rong Jin, Shaili Jain, and Anil Jain. 2013. Inferring users' preferences from crowdsourced pairwise comparisons: A matrix completion approach. In *First AAAI Conference on Human Computation and Crowdsourcing*. 207–215.