Different Font Attributes Influence English Readers' Perception of Font Size Across Age Groups

Yusra Suhail
University of Rhode Island
Kingston, Rhode Island, USA
yusra_suhail@uri.edu

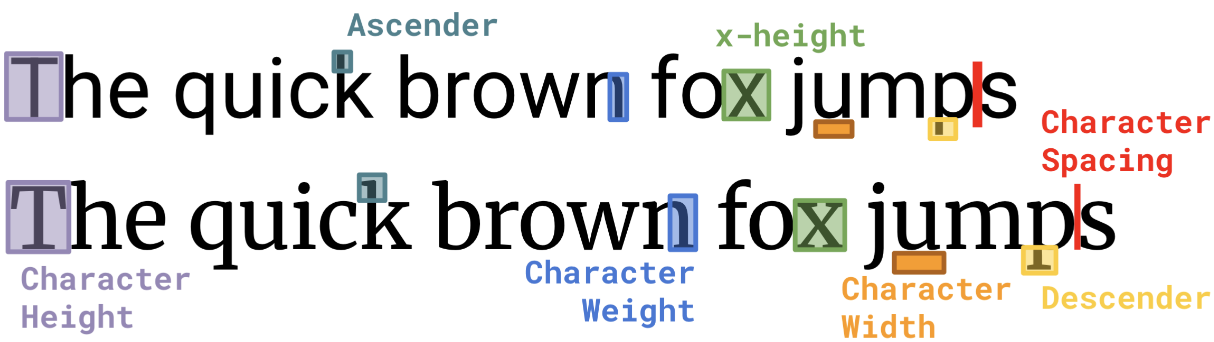
Kushas Khadka
University of Rhode Island
Kingston, Rhode Island, USA
kushas.khadka@uri.edu

Borano Llana
University of Rhode Island
Kingston, Rhode Island, USA
borano.llana@uri.edu

Maedeh Hosseinpour
University of Rhode Island
Kingston, Rhode Island, USA
hosseinpour@uri.edu

Emory Salaun
University of Rhode Island
Kingston, Rhode Island, USA
emory_salaun@uri.edu

Shaun Wallace
University of Rhode Island
Kingston, Rhode Island, USA
shaun.wallace@uri.edu



[h]

Figure 1: This figure shows the seven font attributes on text containing all letters of the Latin alphabet in Roboto Flex (top) and Merriweather (bottom).

Abstract

CSS pixel size is an abstract, device-independent unit that defines the size of elements on digital screens. However, perceived size varies across fonts and display devices. **Font size normalization** aims to remove this inconsistency by matching a font attribute to achieve the same perceived size. Yet, there is no consensus on the best attribute to use for normalization. We aim to personalize this strategy across age groups, as vision changes with age and no single normalization approach works for all fonts [43]. We used a **discrete choice experiment** to study the importance of seven font attributes in perceived size. Character height was the most influential attribute across both age groups; however, age influenced the relative importance of other attributes. Participants 40 or above gave more importance to character weight than character width and spacing, whereas participants under 40 gave the lowest importance to character weight. Our results provide the first evidence for personalizing font size normalization across age groups.

CCS Concepts

• **Human-centered computing** → **Human computer interaction (HCI)**.

Keywords

Readability, Typography, Font Size Normalization, Discrete choice experiments, User Perception, Personalization, Crowdsourcing

1 Introduction and Related Work

Digital reading research and practice rely on a convenient fiction: that a font specified as 12pt, 16px, or 1em corresponds to a stable size that can be compared across devices and typefaces [40]. In 1775, French printer François-Ambroise Didot introduced "point size," a standardized unit to measure printed letters, where

$$1 \text{ point (pt)} = \frac{1}{72} \text{ inch}$$

[31]. The point size of the letter was the height of the metal body on which the letter was cast [34]. With the rise of personal computing and digital displays, font size was measured in physical pixels (px). Early low-resolution screens (≈72 dots per inch (DPI)) approximated

$$1 \text{ pt} = 1 \text{ px}$$

[33]. With the advent of CSS and higher resolution displays, a reference resolution of 96 DPI was defined, establishing CSS pixel as the standard [33]

$$1 \text{ pt} = \frac{4}{3} \text{ px}$$

$$1 \text{ px} = \frac{1}{96} \text{ in.}$$

This calculation, however, is device-independent. With lower or higher resolution screens (different DPI), the same font rendered at the **same pixel size looks smaller or larger on different screens** [33]. Building on this, different devices vary in pixel density. For example, Jung et al. showed that a font height of **8mm** on a

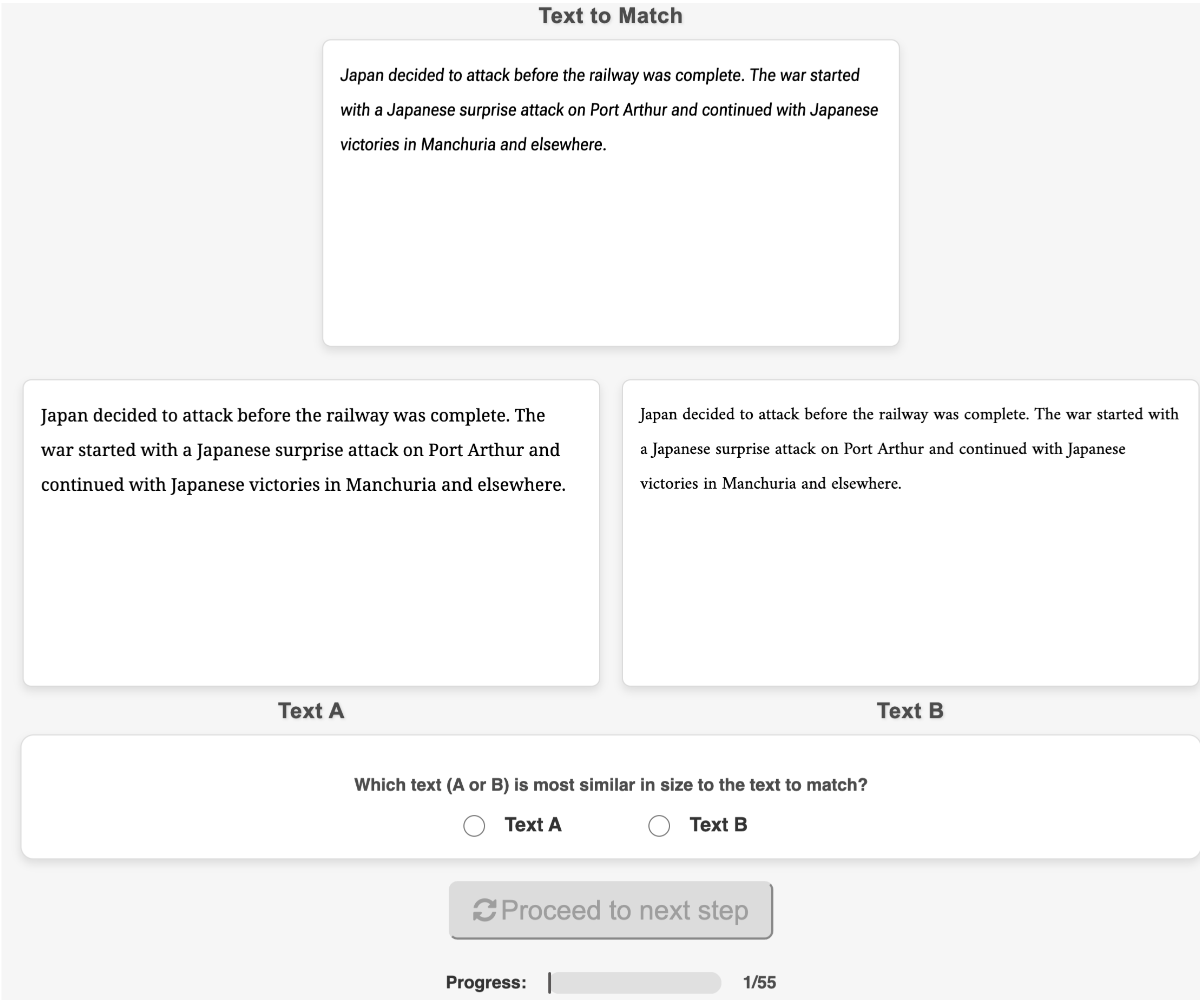


Figure 2: Discrete choice experiment Web UI for Font Size Normalization: The DCE experiment shows a text in the top box rendered in Roboto Flex 16px. The two boxes contain the same text but are rendered in different fonts and sizes.

16-inch display is equivalent to **2mm** on a 4-inch display [21]. Web browsers render fonts in CSS pixels, which cannot be precisely measured for their physical or "real size" [43]. This is the size of the text if someone used a ruler to measure it on a screen. Thus, pixel density across devices influences users' perception of font size. This creates challenges for designers, typographers, and researchers conducting remote readability studies, as fonts are often displayed at unequal sizes [7, 29].

Building on this, comparing different fonts [43], due to differences in font attributes, defined below, some fonts may appear larger or smaller even when displayed at the **same pixel size on the same screen**.

- **Character Height** is the vertical distance from the topmost black pixel of the letter to the last pixel of the letter.
- **x-Height** is the character height of the lowercase letter "x".
- **Character Width** is the horizontal distance of the first black pixel of a letter on the x-axis to the last black pixel.
- **Ascender Height** is the vertical height for the letters t, d, f, h, k, l, b that goes beyond the x-height of that font.
- **Descender Height** is the vertical height for the letters: g, j, p, y that extends below the baseline of a font.
- **Character Spacing** is the horizontal distance between two sequential letters in a word.
- **Character Weight** is the visual density of a character.

We aim to bridge the gap between the font size, measured in pixels and points, and perceived size through **font size normalization**. This technique helps to match the perceived size of **different fonts** even if their default pixel sizes differ.

Font size normalization has been used in past research, where fonts are normalized by matching their x-height, or character height, or character width [2, 43, 46]. However, there is no best candidate font attribute for font size normalization, because *no single* font size normalization strategy is optimal across all fonts [43]. This motivates personalized font size normalization tailored to individual perceptual differences. Research has shown that font size strongly influences perceived size [6, 39]. Legge et al. observed that x-height might influence perceived font size more than average character height [16, 18, 25]. For example, at 12px, Times has an x-height of approximately 2.0mm, while Arial's is closer to 2.5mm [9], making Times appear smaller than Arial. However, Alexander et al. show that these studies did not explicitly investigate users' perception of size itself; instead, they assume that people can accurately perceive font-size differences [1]. Wallace et al. showed that prior research uses fixed font sizes, and should account for readers' perceived font size [9, 43]. This perceived size may also vary with age, as Beier et al found that perceptual abilities decline with growing age [8]. Other studies indicate that age is the most significant factor influencing reading speed and comprehension in neurotypical readers across various reading modes and studies, both in-person and remote [14, 17, 22, 24]. Studies also show that as readers age and their vision weakens, they also experience a decrease in contrast sensitivity and sensitivity to high spatial frequencies[28, 32]. Spatial frequency explains what visual details we can and cannot perceive [30]. While contrast sensitivity is the ability to differentiate between an object and its background, especially when those differences are subtle [23]

Therefore, we explored one core research question through our work:

RQ) How are differences of font size perceived by English readers from different age groups?

2 Method: Discrete choice experiments

Our Institutional Review Board reviewed and approved our methods, procedures, and proposed analysis.

To study users' perceptions of font size across age groups, we designed a ***discrete choice experiment*** (DCE). A DCE is a class of stated preference methods that has commonly been used since the 1980s across multiple disciplines, including health informatics, marketing, economics, and psychology [20, 26, 27], but is less commonly featured in Computer Science research.

In a DCE, each participant sees multiple **choice sets**, each containing a set of mutually exclusive hypothetical alternatives [20, 44]. Participants choose between their preferred alternatives. Each hypothetical alternative is defined by a set of **attributes**, and each attribute can have one or more **levels** [20]. Participants' choices across alternatives quantify the relative importance of each attribute. The **relative importance** shows how influential each attribute is compared to other attributes in a participant's choice.

Table 1 shows an example of an alternative in a discrete choice experiment with three attributes (*font, font size, and if text is bold*). A participant sees multiple of these alternatives, with varying levels for each text. When they make choices between text A & B, it will reveal the relative importance they give to each attribute when making a choice.

In our discrete choice experiment, we showed participants two texts (A & B) (*alternatives*) containing identical content but rendered at different font size combinations (See Figure 2). They were asked to choose the one they believed best matched the reference text (called the text to match in our study) in size. Each font size combination produced different measurements for each of the seven font attributes shown in Figure 1 (which were our DCE attributes).

Table 1: Example of a discrete choice experiment (DCE). Choices could be text A or B. The attributes for a DCE are on the left, while the levels or variations in the attributes are listed in the columns under Text A and B. Participants are given a choice (option A or B) based on a question. For example, this choice set below might be asking a designer to select which font (text design) they would prefer to read in.

	Text A	**Text B**
Font Size	18px	16px
Line Spacing	1.0	1.25
Bold	Yes	No

We used canvas text metrics to measure these attributes by rendering all Latin alphabets at that font-size combination on an HTML5 canvas and averaging across each combination. We then computed the difference between the font attribute measurements of the text the participant selected and the text to match. These differences were treated as the levels of each attribute. All measurements were recorded in centimeters, in accordance with the user's screen specifications, to control for pixel size inconsistencies across devices. This method directly improves on prior work by integrating real-world units in size across devices [14].

We approximated character weight as the visual density of the character and calculated it using Shannon's entropy of the alpha pixels (which measure transparency of an image) of the character's image rendered on canvas. Shannon's entropy measures the randomness in a signal or an image [37, 42]. In image processing, a detailed image has higher entropy than a uniform image. In our case, it measures the distribution of black pixels, where higher entropy indicates a higher visual density.

For this experiment, we rendered the text to match in Roboto Flex at 16px. Roboto Flex was chosen because it is a variable font with five user axes and seven parametric axes. Our study system used axes that allowed us to vary the text to match in character weight, character spacing, ascender height, descender height, x-height, and character width. We did not use the numeric parametric axis. We created **675** unique combinations of Roboto Flex, each having a random value for all of its *11 axes*.

We consulted typographers to select font sizes and families for our discrete choice experiment. We selected three font sizes (14px, 16px, and 18px). These sizes were chosen in light of previous studies and for the recommendation to use font sizes above 14px [5, 10, 38, 43]. We selected the variable font Roboto Flex as the control font, and then we consulted with typographers to ensure our 8 fonts to compare covered a wide range of font attributes (i.e, tall ascenders and descenders or a small x-height):

- **Poppins**: features tall x-height.
- **Georgia**: a serif font with tall ascenders.
- **Merriweather**: has a large x-height and has slightly condensed letter forms (reduced horizontal width).
- **Noto Serif**: smaller stroke widths than Merriweather.
- **Amiri**: has tall ascenders.

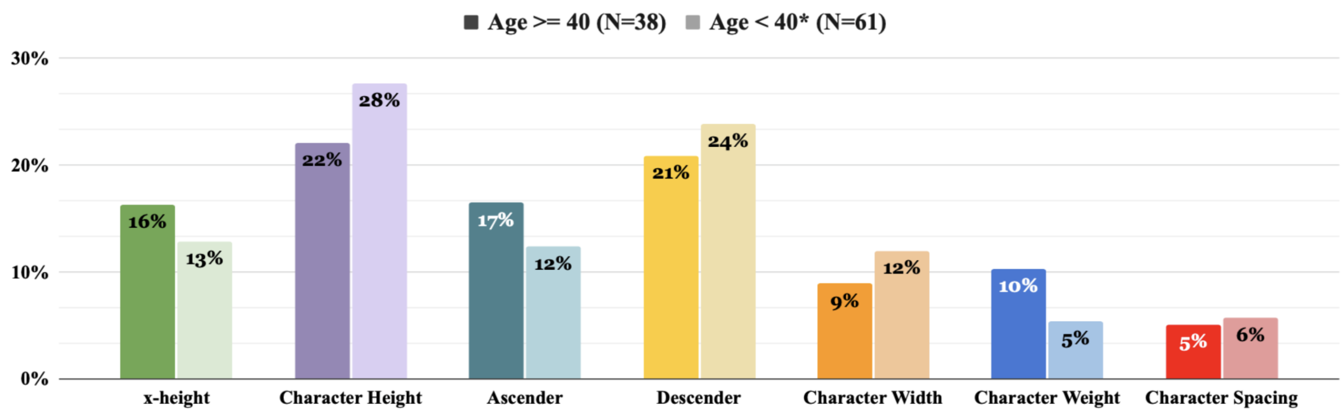


Figure 3: Relative Importance of Font Attributes for 40 or Above and Under 40 (ages 18-39) Native English Readers: Dark shade represents 40 or above, and light shade represents under 40 participants.

- **Noto Naskh Arabic**: is commonly used in writing administrative documents and for transcribing books, including the Quran, because of its easy legibility.
- **Lateef**: has a small x-height.
- **Lemonada**: has a tall x-height and short ascenders.

With three font sizes and eight fonts, we created **276** unique font pairs. Each participant was shown **50** choice sets, and each choice set displayed one of the randomly picked 276 font pairs for each text A & text B and one randomly selected Roboto Flex variation for text to match. Thus, in our study, participants did not observe attributes or their levels directly (see Figure 2).

3 Results and Discussion

We conducted a conjoint analysis using data collected from our DCE. A conjoint analysis helps explain the relative importance of attributes for users based on their responses of stated preferences [15, 36, 44].

We performed this analysis using an Ordinary Least Squares (OLS) regression model per recommendations from prior research[19, 41]. Prior research by Simon et al. recommends OLS to analyze choice data because 1) it provides a simple, interpretable way to estimate the relative importance of a discrete choice experiment's attributes; 2) our N per group is smaller; and 3) the risk of overfitting in low-dimensional data is less than in more complex models [41]. Our study also follows a randomized conjoint DCE framework, where attribute levels are randomized per choice set. Hainmueller et al. show that Ordinary Least Squares can be used in such designs to estimate the Average Marginal Component Effects (AMCEs)[19]. We estimated the effect of each attribute on the probability that a text is selected using OLS.

3.1 Font Perception Across all Groups

Overall, our results show that character height was the most influential attribute among all participants (see Figure 3). This result mirrors prior research by Wallace et al., which found that character height was the most frequently selected option for font size normalization among x-height, character height, and character width [43]. However, our study compared seven font attributes, providing new evidence for what affects readers' font size perception.

The second most important attribute for all participants was descender height. However, for our participants over 40, the relative importance of character height and descender was similar (22% vs 21%). Our results suggest that perception of font size is shaped by font attributes beyond character height, x-height, and width, the primary focus of prior studies [43].

Our results contrast prior research that identified x-height as the most important font attribute in determining perceived font size [12, 16, 25, 35, 45]. In our study, x-height is the fourth most important font attribute in font size normalization for readers 40 or older, and the third most important for readers under 40.

3.2 Weight was more Important for Participants Above 40 than for Participants Aged 18–39

To analyze the effect of age-related visual changes on font perception, we recruited 99 paid crowdworkers using Prolific and divided them into two groups: 61 participants (ages 18-39) and 38 participants (40 and older). Among the 18-39 age group, 22 identified as male, 37 as female, and 2 as other. Among the 40 and older age group, 19 identified as male and 19 as female.

We chose 40 as the age to split participants based on recommendations from the American Optometric Association. According to the American Optometric Association, many adults begin to experience *presbyopia* in their early to mid-40s: a condition that impairs near vision and can affect perceived font size and clarity [3, 4, 13]. To account for these perceptual differences, we analyzed these groups separately.

Our results show that, for participants 40 or older, character weight was twice as important as for participants under 40 (see Figure 3). Prior research has shown that contrast sensitivity decreases with age [28]. This measure indicates that participants aged 40 and older were more sensitive to changes in font weight than those under 40.

3.3 Towards Future Work on the User Perception of Character Details and Spatial Density across Age Groups

Prior research shows that older adults may have weaker or variable eyesight, making it challenging to identify character (i.e., glyph) details [14]. According to the Figure 3, our results show that for participants under 40, descender height was twice as important as ascender height (24% vs 12%). Whereas, for participants over 40, descender height was 21% as important as ascender height (21% vs 17%). Furthermore, for participants under 40, the importance of x-height was similar to character width (13% vs 12%). Whereas, for participants over 40, x-height was 56% more important than character width (16% vs 9%).

These results may also be explained by the prior research on spatial frequency. Readers often perceive individual character or glyph details in a box surrounding each character. Thus, Mugikura et al. noted that one letter contains various spatial frequencies, where the main letter requires mid-level spatial frequency, while the corners and sharp edges of the letter require high spatial frequency [30]. High spatial frequencies correspond to tiny details in images, while low spatial frequencies correspond to large, diffused features [11]. Prior research shows that sensitivity to high and mid-level spatial frequency decreases with age [28, 32].

We posit that ascender and descender are likely to have higher spatial frequencies than x-height. Visual information for a character's ascender and descender resides near the outer limits. Also, character width and weight consist of low spatial frequencies that correspond to denser visual features [11]. Therefore, the differences we observe based on age might be attributed to reader's perception of characters across this "perceptual box" the surrounds each glyph.

Therefore, this helps explain why the relative importance of font attributes with low to mid spatial frequencies trend higher for our participants aged 40 and above. In contrast, participants under 40 appeared to rely more on high spatial frequency cues, as descender height was twice as important as ascender height and x-height. We call for future work to examine font-size perception and normalization by analyzing the spatial frequencies of characters (i.e., glyphs) across font families.

3.4 Qualitative Results on How Readers Perceived Font Size and Made their Choices

After the main study, our participants completed a post survey. We performed thematic analysis to identify common themes in participants' responses to an open-ended question: *"What strategies did you use to choose which font was close in size to the control font?"*. We identified three codes: spacing (between lines and words), word-level (comparing whole words), and character-level (comparing individual characters); see Table 2.

Our qualitative results showed that both age groups valued word-level features much more highly than the character-level features. However, most existing font features are designed at the character level. Our findings highlight the need to design font attributes at the word level, such as word-level spacing.

Table 2: Thematic Analysis: codes on how participants matched fonts. *Spacing is line spacing and line breaks.

Group	Word-Level	Character-Level	Spacing*
Age $\geq$ 40	42%	5%	47%
Age < 40	34%	5%	39%

4 Conclusion and Future Work

This work presents the initial effort to personalize font-size normalization across age groups. To achieve this, we conducted a **discrete choice experiment** to study the relative importance of seven font attributes in readers' perception. Our work revealed that as readers age, their perception of font size also changes. For example, when matching two fonts, participants 40 or above placed greater importance on character weight than on character width and spacing, whereas those under 40 placed the least importance on character weight. This work demonstrates the need for demographic-sensitive design techniques when designing and normalizing fonts.

Future work should explore how readers' perceptions of font size vary among those with different vision conditions, such as myopia (nearsightedness), hyperopia (farsightedness), and astigmatism, as well as among those with varying contrast sensitivity.

5 Acknowledgments

We want to thank the organizers and participants of the CHI 2026 Science and Technology for Augmenting Reading (STAR) workshop. This paper was presented at the workshop. We also want to thank Sam Berlow for his advice and guidance throughout this work.

References

[1] Eric Alexander, Chih-Ching Chang, Mariana Shimabukuro, Steven Franconeri, Christopher Collins, and Michael Gleicher. 2017. Perceptual biases in font size as a data encoding. *IEEE transactions on visualization and computer graphics* 24, 8 (2017), 2397–2410.

[2] Mrouj Almuhajri. 2013. *Arabic E-Reading: Studies on Legibility and Readability for Personal Digital Assistants.* Ph. D. Dissertation. Concordia University.

[3] American Optometric Association. [n. d.]. Adult Vision: 41 to 60 Years of Age. https://www.aoa.org/healthy-eyes/eye-health-for-life/adult-vision-41-to-60-years-of-age

[4] Masahiko Ayaki, Akiko Hanyuda, and Kazuno Negishi. 2025. Presbyopia and associated factors specific to age groups. *Clinical and Experimental Optometry* (2025), 1–6.

[5] 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.

[6] Scott Bateman, Carl Gutwin, and Miguel Nacenta. 2008. Seeing things in the clouds: the effect of visual features on tag cloud selections. In *Proceedings of the nineteenth ACM conference on Hypertext and hypermedia*. 193–202.

[7] Sofie Beier, Sam Berlow, Esat Boucaud, Zoya Bylinskii, Tianyuan Cai, Jenae Cohn, Kathy Crowley, Stephanie L Day, Tilman Dingler, Jonathan Dobres, et al. 2022. Readability research: An interdisciplinary approach. *Foundations and Trends® in Human–Computer Interaction* 16, 4 (2022), 214–324.

[8] Sofie Beier and Chiron Oderkerk. 2019. Visible Language - The effect of age and font on reading ability. *Visible Language* 53, 3 (Dec 2019), 50–69. doi:10.34314/vl.v53i3.4654

[9] 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.

[10] 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), 156–179.

[11] Erik Blaser. [n. d.]. Spatial vision. https://psych.umb.edu/blaser/blaserWebsite/Psych_355_%28Perception%29/Entries/1000/1/1_18._Spatial_vision.html

[12] 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*. 87–94.

[13] Kierstan Boyd. 2024. What is presbyopia? https://www.aao.org/eye-health/diseases/what-is-presbyopia

[14] Tianyuan Cai, Shaun Wallace, Tina Rezvanian, Jonathan Dobres, Bernard Kerr, Samuel Berlow, Jeff Huang, Ben D Sawyer, and Zoya Bylinskii. 2022. Personalized font recommendations: Combining ml and typographic guidelines to optimize readability. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference*. 1–25.

[15] Conjointly. [n. d.]. How to interpret Partworth Utilities. https://conjointly.com/guides/how-to-interpret-partworth-utilities/

[16] Bart Cooreman and Sofie Beier. 2025. Research highlight: How important is X-height for font legibility? https://readabilitymatters.org/articles/research-highlight-how-important-is-x-height-for-font-legibility

[17] 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.

[18] Google Fonts. 2023. Exploring x-height & the em square. https://fonts.google.com/knowledge/choosing_type/exploring_x_height_the_em_square Accessed: 2025-01-03.

[19] Jens Hainmueller, Daniel J Hopkins, and Teppei Yamamoto. 2014. Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments. *Political analysis* 22, 1 (2014), 1–30.

[20] David Hoyos. 2010. The state of the art of environmental valuation with discrete choice experiments. *Ecological Economics* 69, 8 (2010), 1595–1603. doi:10.1016/j.ecolecon.2010.04.011

[21] Sungwook Jung, Youngjae Lim, Eui S Jung, and Jean Jung. 2010. Cultural differences in display design for mobile web-browsing tasks. In *2010 4th International Conference on Multimedia and Ubiquitous Engineering*. IEEE, 1–6.

[22] Tami Katzir, Shirley Hershko, and Vered Halamish. 2013. The effect of font size on reading comprehension on second and fifth grade children: bigger is not always better. *PloS one* 8, 9 (2013), e74061.

[23] Kirandeep Kaur and Bharat Gurnani. 2022. Contrast sensitivity. (2022).

[24] Susan Kemper and Joan McDowd. 2006. Eye movements of young and older adults while reading with distraction. *Psychology and aging* 21, 1 (2006), 32.

[25] Gordon E Legge and Charles A Bigelow. 2011. Does print size matter for reading? A review of findings from vision science and typography. *Journal of vision* 11, 5 (2011), 8–8.

[26] Jordan J Louviere and David A Hensher. 1982. On the design and analysis of simulated choice or allocation experiments in travel choice modelling. *Transportation research record* 890, 1 (1982), 11–17.

[27] Jordan J Louviere and George Woodworth. 1983. Design and analysis of simulated consumer choice or allocation experiments: an approach based on aggregate data. *Journal of marketing research* 20, 4 (1983), 350–367.

[28] Clare McGrath and JD Morrison. 1981. The effects of age on spatial frequency perception in human subjects. *Quarterly Journal of Experimental Physiology: Translation and Integration* 66, 3 (1981), 253–261.

[29] 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, ACM, New York, NY, USA, 285–296.

[30] Shoko Mugikura. 2014. How spatial frequencies affect the way we perceive fonts. https://blog.justanotherfoundry.com/2014/06/how-spatial-frequencies-affect-the-way-we-perceive-fonts/

[31] Jeremy M. Norman. 2025. François-Ambroise Didot Revises Fournier's Point System for Typographic Units. https://www.historyofinformation.com/detail.php

[32] Cynthia Owsley, Robert Sekuler, and Dennis Siemsen. 1983. Contrast sensitivity throughout adulthood. *Vision research* 23, 7 (1983), 689–699.

[33] Steven Pemberton. 2002. A pixel is not a point. *interactions* 9, 5 (2002), 64.

[34] thomas Phinney. 2025. Font point size and the EM Square: Not what people think. https://www.thomasphinney.com/2011/03/point-size/

[35] Eustace Christopher Poulton. 1965. Letter differentiation and rate of comprehension in reading. *Journal of Applied Psychology* 49, 5 (1965), 358.

[36] Vithala R Rao. 2010. Conjoint analysis. *Wiley International Encyclopedia of Marketing* (2010).

[37] QR Razlighi and Nasser Kehtarnavaz. 2009. A comparison study of image spatial entropy. In *Visual Communications and Image Processing 2009*, Vol. 7257. SPIE, 615–624.

[38] 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*. 3637–3648.

[39] Anna W Rivadeneira, Daniel M Gruen, Michael J Muller, and David R Millen. 2007. Getting our head in the clouds: toward evaluation studies of tagclouds. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 995–998.

[40] Nick Sherman. 2013. Responsive typography is a physical discipline, but your computer doesn't know it (yet). https://alistapart.com/column/responsive-typography-is-a-physical-discipline/

[41] Gyorgy Simon and Constantin Aliferis. 2024. An Appraisal and Operating Characteristics of Major ML Methods Applicable in Healthcare and Health Science. *Artificial Intelligence and Machine Learning in Health Care and Medical Sciences: Best Practices and Pitfalls* (2024), 95–195.

[42] Du-Yih Tsai, Yongbum Lee, and Eri Matsuyama. 2008. Information entropy measure for evaluation of image quality. *Journal of digital imaging* 21, 3 (2008), 338–347.

[43] Shaun Wallace, Zoya Bylinskii, Jonathan Dobres, Bernard Kerr, Sam Berlow, Rick Treitman, Nirmal Kumawat, Kathleen Arpin, Dave B Miller, Jeff Huang, et al. 2022. Towards individuated reading experiences: Different fonts increase reading speed for different individuals. *ACM Transactions on Computer-Human Interaction (TOCHI)* 29, 4 (2022), 1–56.

[44] Shaun Wallace, Talie Massachi, Jiaqi Su, Dave B Miller, and Jeff Huang. 2025. Towards Fair and Equitable Incentives to Motivate Paid and Unpaid Crowd Contributions. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3706598.3714195

[45] Arnold Wilkins, Roanna Cleave, Nicola Grayson, and Louise Wilson. 2009. Typography for children may be inappropriately designed. *Journal of Research in Reading* 32, 4 (2009), 402–412.

[46] Bin Zhang. 2011. *Discovering Legible And Readable Chinese Typefaces For Reading Digital Documents*. Master's thesis. Concordia University. https://spectrum.library.concordia.ca/35807/1/Zhang_MCompSc_F2011.pdf