Typefaces Are All Almost the Same

Non-systematic Review of Scientific Findings on the Impact of Typographic Features on Reading Performance

Luciano Perondi
Università Iuav di Venezia
lperondi@iuav.it
ORCID 0000-0002-7198-4030

Abstract
The interest in discovering the limits of legibility has not ceased to interest those who use letters in typography and, to a certain extent, those who study the reading process. Sometimes the debate on the subject is marred by sensationalism.

The research does not aim to identify a particular typeface and to promote it as "best for legibility". The research aims to identify what are the possible reasons that make one typeface better than another and to make this knowledge available to the community, so that the typefaces can guarantee better reading performance, if this is possible.

Keywords
Legibility
Reading process
Typeface design
Readability

Introduction

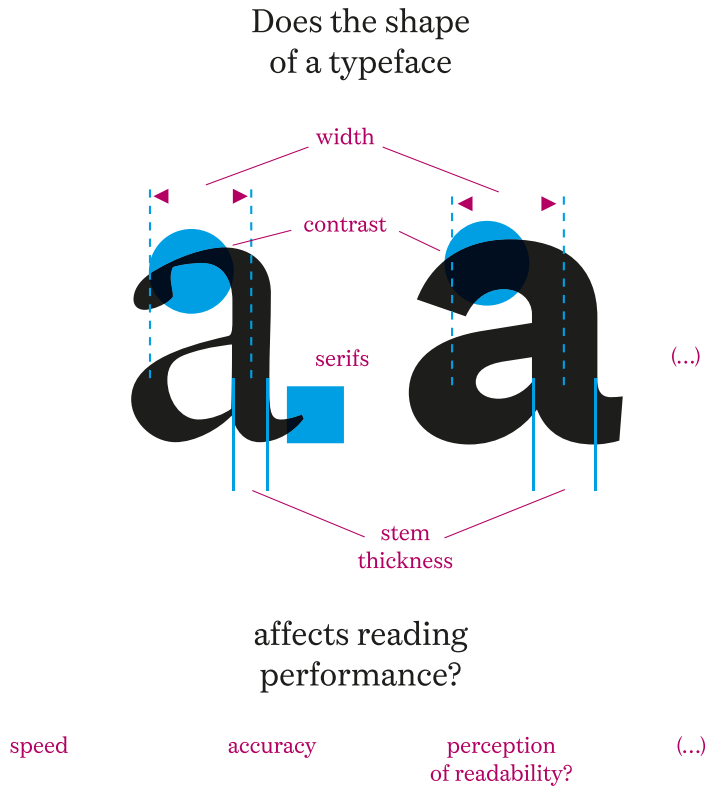
Studies on reading have an ancient origin and the interest for discovering the limits of the legibility of letters has not ceased to interest those who use letters (typographers and graphic designers) and those who study the phenomenon of reading.

In the field of graphic design, professionals and scholars ought to keep in mind the effects of a particular typographic feature on reading, in order to operate free from subjective prejudices while choosing or designing a typeface for a text. Indeed, in several experiments (Bernard et al., 2001; Darroch et al., 2005; Larson & Carter, 2016) the authors found that the "perception of legibility" is not a good predictor of actual reading performance.

The current research aims to identify what are the features that make one typeface better than another in a given context and to make this knowledge available to the community.

I reviewed a series of scientific articles discussing the impact of typeface features on reading performance.

I will focus on the features of the letters and therefore I'll exclude the impact of other variables that are related to the typesetting, such as the spaces between letters and words, the line length, and the line spacing.



diid No. 74 — 2021
Doi: 10.30682/diid7421i

Legibility and the Assessment of the Reading Performance

Researching on legibility certainly implies referring to what is known about the reading process, a complex and multifaceted phenomenon, which involves physical, physiological and perceptive aspects (Price, 2012; Carreiras et al., 2014; Rastle, 2016; Franceschini et al., 2017).

Charles Bigelow (2019) and Tarasov et al., (2015) provide a detailed state of the art on research related to legibility and its ultra-centennial history.

I will start from two issues:

1 the construct of legibility and
2 the method used to study the reading performance.

Literature consistently distinguishes between "legibility", i.e., the recognition of letters and words, and "readability", i.e., the reading comprehension (Bollini et al., 2020; Rand, 2002). In this essay I will focus mostly on legibility and on the measure of reading performance.

Studying reading performance means to measure how one or more independent variables (e.g. a formal feature such as the shape of the serif or the thickness of the strokes) is related to one or more dependent variables (e.g. speed of word recognition (Hughes & Wilkins, 2000; 2002), accuracy, text acquisition (Sheedy et al., 2005), number of errors in reading, comprehension (Dyson & Kipping, 1998), visual searching (Huang et al., 2009) etc.).

If the correlation between the two variables is statistically significant, it can therefore be generalized that the given variable(s) (e.g. a visual feature of a type) determines an effect on the reading performances.

Some of these performances sometimes turn out to be predictors of other performances. Tarasov et al. (2015) date back to 1885 (Cattell, 1885) the evidence that reading speed is a good predictor of other variables connected with legibility.

The difficulty of the process therefore lies in isolating the specific features of the typefaces, measuring them and seeing if these variations have a significant impact on a particular reading performance.

Measuring the Typefaces

Two problems arise while isolating the features of a typeface:

1 the former concerns the nature of measurements in typography: a typeface is a set of distinct glyphs which are vaguely regular in their shapes and only for some features (size, thickness of the stems etc.);
2 the latter concerns the generalizability of the results: often a broad variety of typefaces with noticeably different shapes and measures are used in experiments. For this reason several features (and therefore typographic variables) are often aggregated; e.g. Courier and Times have been

compared in some experiments (Mansfield et al., 1996; Gasser et al., 2005), but the two typefaces differ in several features such as typographical contrast, letter width, proportional or monospaced spacing.

Even if the experiment investigates the combined effect of an aggregation of variables, the aim is preferably to understand which features in what specific context make a particular typeface perform better than another. The variables must be under control, in order to know exactly what the experiment is measuring.

Studying the effect of aggregates of variables, however, can prove useful in a process of progressive identification of which are the features that have the greatest impact on the reading, or if there are particular effects combined between two or more features.

In a recent work in which I took part (Galliussi et al., 2020) we decided to study together three distinct visual features of the typeface (ascenders height, asymmetry of shapes and presence of "dyslexia friendly" letters). Even though it was possible to isolate the variables, we preferred to consider them together in order to investigate how their effects interacted.

In order to make shape feature comparisons without font size affecting performance, fonts must be used at similar apparent size.

Legge and Bigelow (2011) provide reasons to support the fact that x-height provides a measurement of the size of the glyphs that allows a better comparison between different typefaces (regardless the influence of the shapes of a typeface on the perceived size).

Indeed, the measure of x-height instead of the body is now also used in legislative contexts (for example the Regulation (EU) No 1169/2011).

However, the measurement of the x-height is not sufficient to define the size of a text. It is necessary to define the reading distance, in order to understand what is the visual angle subtended by this measure (Bigelow, 2019).

On the basis of a large amount of empirical evidence, Legge (2006; Bigelow, 2019) indicates that the critical size of x-height corresponds to 1.5 mm at 40 cm of distance, which means 0.21 degrees of visual angle. He also observes that the reading speed remains more or less constant between 0.2 and 2 degrees, a measure beyond which (2006; Legge & Bigelow, 2011) he indicates that the effects of parafoveal benefits are lost, as already pointed out by Tinker (1963).

Rubin et al. (2006) note that below what they define "Critical Print Size" (CPS) the reading speed of a text decreases significantly. He also notes that increasing the font size significantly increases the portion of the population that can read fluently.

Although the performances within the range of 0.2-2 degrees may vary significantly (Chandler, 2001), the difference in performance in the range indicated is very low and not very relevant in practice.

diid No. 74 — 2021
Doi: 10.30682/diid7421i

In this context is relevant the Beier and Oderkerk experiment (2019). This experiment highlights a "drop-off at the lightest and the heaviest extremes at all tested font sizes", which suggests that the feature 'thickness of the strokes' affect the performance of reading in a similar way of size: a large central area, from light to bold, where the reading performances are more or less aligned. The experiment "provides evidence that bolder fonts are less effective at improving recognition at larger visual angles" while they are more effective in improving recognition in small size. The experiment, however, measured the accuracy in the recognition of single letters, a very precise task in terms of typographical features, but rather distant from an ecological reading task.

However, several experiments do not equalize font x-height.

For example, in a 2013 experiment Bachmann (2013) compares two typefaces by submitting a text to normotypic and dyslexic readers. The two fonts are set at the same body size, but the two typefaces have two significantly different x-heights, different line spacing, different space between words and between letters.

Although a significant effect has been measured, it is not possible to distinguish the effect of the shape features of the fonts from that of the size and spacing.

This does not affect the scientific validity of an experiment, but it cannot lead to generalizations about the effects of a specific feature of a typeface, above all because in literature there is evidence of the effect of size and spacing on reading performance.

Similarly, Larson and Carter (2016) found that the Sitka display version was less performing at both small and large sizes than the small body version, but again the typographic features involved concerned the width of the letters, the typographic contrast, the x-height, the length of the serifs.

Serif vs. Sans and Other Features

The most recurring comparison in literature is that between serif and sans typefaces.

A large number of studies (De Lange et al., 1993; Chandler, 2001; Shaikh, 2007; Arditi & Cho, 2005) concluded that there are no differences in legibility between serifs and sans.

Chandler (2001), compares Palatine and Helvetica at different sizes, but does not detect significant differences between one typeface and the other.

The same notes Rubin et al. (2008) who compare different typefaces at different sizes: "There was also a significant effect of typeface with TPC being read about 8 words min^{-1} faster, on average, than the other fonts (159 words min^{-1} for TPC vs 151 words min^{-1} for the other fonts, $p < 0.0001$). However, fonts of the same nominal point size were not equivalent in actual size. When adjusted for the actual horizontal and vertical space occupied, the advantage of TPC was eliminated".

A trend that has had extensive development in recent years is that of the "Dyslexia Friendly" typeface, a research area to which I personally contributed (Perondi et al., 2017; Galliussi et al., 2020). On this subject, in previous articles we have reported a series of ev-

idences (Marinus et al., 2016; Duranovic et al., 2018; Kuster et al., 2017; Wery et al., 2017; Bollini et al., 2020) who compared Times New Roman, Simoncini Garamond, Verdana, Futura, Comic Sans, OpenDyslexic, EasyReading, Biancoenero in legibility and readability tasks) in which the differences found in the reading performances were not statistically significant, like in the experiments we conducted (Perondi et al., 2017; Galliussi et al., 2020).

It can be concluded that there is no evidence "of a facilitatory effect of the DF dedicated graphic features embedded in the letterform" (Galliussi et al., 2020). These results concerned readers from different countries where Latin characters are used, so the results seem to be language independent.

Measuring the Effects

It may happen that the effects on performance are not consistent, for example a typeface can work well in terms of aiding the speed of recognition of letters in the readers, but being weak in aiding accuracy or lessening reading fatigue. It is therefore not obvious to obtain an overall measure of the readability of a typeface.

Bigelow (2019) reports a study by Mansfield, Legge, and Bane (1996), in which there is evidence that Times Roman appears to be "5% faster than Courier Bold at moderate print sizes (above critical print sizes for the respective fonts), but read Courier Bold up to twice as fast as Times at tiny sizes (below critical print sizes)" and states that different typeface features can affect reading, but it is difficult to isolate the effect of these features. Tarita-Nistor et al. (2013) found in readers with macular degeneration that at dimensions close to their visual acuity in reading (reading acuity) the Courier is more efficient than the Arial for what concerns the reading acuity but not for what concerns the reading speed.

And they point out "This is contrary to the advice given by agencies for the blind".

Lund (1999) is quite critical about experimenting with typeface features, underlining that differences between one typeface and another can also be detected in terms of a specific performance, but in practice these differences are so small or specific as to be ephemeral, and the scientific evidence in literature seems to confirm this observation.

In general, I observe that there is difficulty in defining experimental tasks close to the reading experience, so the measurements often have not a strong ecological validity. On the other hand, if the measurements are ecological, there are often too many variables that can affect the reading process; as noted by Beier and Oderkerk (2019) "[...] when fonts of different families are tested, the fonts vary on several parameters, and it becomes difficult to identify which features have an effect on the results".

Luciano Perondi
Typeface and graphic designer, Researcher, Associate Professor of Design at the Università Iuav di Venezia, his interests are focused on visual studies, writing, typography and information design.

Reader Subjective Opinion

In an experiment to which I contributed (Perondi et al., 2017), we submitted to the subjects the request to sort in order of "readability" eight sheets with the same text composed with eight different typefaces. We recognized some patterns: one particular typeface was regularly recognized as "less legible" than the others. When we tested the actual users' reading performance, however, this typeface did not decrease reading performance compared to the others.

This evidence indicates that reader's opinion is not a good predictor of reading performance, in accordance with much literature (e.g. Bernard et al., 2001; Darroch et al., 2005), in which subjects underestimate their ability to read small-sized texts and tend to prefer much larger fonts than they read efficiently.

Similarly, in the Sitka project (Larson & Carter, 2016) the experimental evidence showed a contradiction between actual reading performance and readers' and experts' opinion on the legibility and readability.

Very relevant is the fact that Huang, Rau and Liu (2009) in an experiment on the optimal size of Chinese typefaces on mobile screens found that also for Chinese characters the "Subjective measures were found to be more sensitive than objective measures".

This highlights a very important aspect: readability can be understood as the expressive quality (Sinico, 2015) of a typeface.

Conclusions

Observing the literature analyzed, the specific variables relating to the shape of the typefaces appear to have a non-significant impact or at best minimal or very specific. The size has a greater impact, even if the range within the two critical thresholds (minimum critical print size and maximum critical size) has little marked differences.

It can therefore be induced that research has identified few significant results of low practical utility. Furthermore, the research is complicated by some distance between the setting of the experimental tasks and the ecological reading process.

What emerges from the evidence is that there is ample freedom in the design of the features of the typefaces, because these do not seem to show particular impacts on the reading performance.

On the other hand, the subject of the expressive qualities of typefaces (Velasco & Spence, 2018; 2019) is promising.

Readers' intersubjective attitude towards the expressive qualities of the text can affect the perception of the text itself and what the text communicates in terms of readability.

In this sense, identifying type features that increase the perceived readability of a text can be functional to reduce the threshold of access to the text itself: using a typeface that appears more readable, can entice the reader to face a text, and in the same time that typeface allows the designer to trust in the fact that the reading performance will tend to be in line with that of most of the more popular typefaces.

References

Arditi, A., & Cho, J. (2005). Serifs and font legibility. *Vision research, 45*(23), 2926-2933. https://doi.org/10.1016/j.visres.2005.06.013

Bachmann, C. (2013). Può un font essere uno strumento compensativo per i lettori con dislessia. Gradimento e prestazione nella lettura in Times New Roman e in EasyReading® di alunni dislessici e normolettori della classe quarta primaria. *Dislessia, 10*(2), 243-262. https://www.easyreading.it/wp-content/uploads/2017/03/bachmann_-scientific_study.pdf

Beier, S., & Oderkerk, C.A. (2019). Smaller visual angles show greater benefit of letter boldness than larger visual angles. *Acta psychologica, 199*, 102904. https://doi.org/10.1016/j.actpsy.2019.102904

Beier, S., & Oderkerk, C.A. (2019). The effect of age and font on reading ability. *Visible Language, 53*(3), 50-68. https://www.proquest.com/docview/2330018843?pq-origsite=gscholar&fromopenview=true

Bernard, M., Liao, C.H., & Mills, M. (2001, March). 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* (pp. 175-176). https://doi.org/10.1145/634067.634173

Bigelow, C. (2019). Typeface features and legibility research. *Vision research, 165*, 162-172. https://doi.org/10.1016/j.visres.2019.05.003

Bollini, L., Brambilla, V., & Stucchi, N.A. (2020). The impact of high-legibility fonts in dyslexic people reading experience: an experimental evaluation. In *Proceedings of ICERI2020 Conference* (pp. 778-788). IATED. 10.21125/iceri.2020.0231

Carreiras, M., Armstrong, B.C., Perea, M., & Frost, R. (2014). The what, when, where, and how of visual word recognition. *Trends in cognitive sciences, 18*(2), 90-98. https://doi.org/10.1016/j.tics.2013.11.005

Cattell, J.M. (1885) The inertia of the eye and the brain. *Brain, 8*, 295-312.

Chandler, S.B. (2001). *Comparing the legibility and comprehension of type size, font selection, and rendering technology of onscreen type*. Virginia Polytechnic Institute and State University. https://www.proquest.com/docview/304729532?pq-origsite=gscholar&fromopenview=true

Darroch, I., Goodman, J., Brewster, S., & Gray, P. (2005, September). The effect of age and font size on reading text on handheld computers. In *IFIP conference on human-computer interaction* (pp. 253-266). Springer. https://link.springer.com/chapter/10.1007/11555261_23

De Lange, R.W., Esterhuizen, H.L., & Beatty, D. (1993). Performance differences between Times and Helvetica in a reading task. *Electronic Publishing, 6*(3), 241-248. http://cajun.cs.nott.ac.uk/compsci/epo/papers/volume6/issue3/rudi.pdf

Duranovic, M., Senka, S., & Babic-Gavric, B. (2018). Influence of increased letter spacing and font type on the reading ability of dyslexic children. *Annals of dyslexia, 68*(3), 218-228. https://link.springer.com/article/10.1007/s11881-018-0164-z

Dyson, M.C., & Kipping, G.J. (1998). The effects of line length and method of movement on patterns of reading from screen. *Visible language, 32*(2), 150-181. https://www.proquest.com/openview/49f4fd2296eb9ecef039e3cd3921fe0b/1?pq-origsite=gscholar&cbl=1821103

Franceschini, S., Bertoni, S., Gianesini, T., Gori, S., & Facoetti, A. (2017). A different vision of dyslexia: Local precedence on global perception. *Scientific reports, 7*(1), 1-10. https://www.nature.com/articles/s41598-017-17626-1

Galliussi, J., Perondi, L., Chia, G., Gerbino, W., & Bernardis, P. (2020). Inter-letter spacing, inter-word spacing, and font with dyslexia-friendly features: testing text readability in people with and without dyslexia. *Annals of dyslexia, 70*(1), 141-152. https://dx.doi.org/10.1007%2Fs11881-020-00194-x

Gasser, M., Boeke, J., Haffernan, M., & Tan, R. (2005). The Influence of Font Type on Information Recall. *North American Journal of Psychology, 7*(2). https://www.researchgate.net/profile/Michael-Gasser/publication/237229931_The_Influence_of_Font_Type_on_Information_Recall/links/0046352ffbb8566bd4000000/The-Influence-of-Font-Type-on-Information-Recall.pdf

Hashim, A.A., & Majid, M. A. (2015). Effects of video display terminal resolutions to the legibility of text on a web page. *International Journal of Software Engineering and Computer Systems, 1*(1), 131-157. https://ijsecs.ump.edu.my/images/archive/vol1/11Affendi_IJSECS.pdf

Huang, D.L., Rau, P.L.P., & Liu, Y. (2009). Effects of font size, display resolution and task type on reading Chinese fonts from mobile devices. *International Journal of Industrial Ergonomics, 39*(1), 81-89. https://doi.org/10.1016/j.ergon.2008.09.004

Hughes, L., & Wilkins, A. (2000). Typography in children's reading schemes may be suboptimal: Evidence from measures of reading rate. *Journal of Research in Reading, 23*(3), 314-324. https://doi.org/10.1111/1467-9817.00126

Hughes, L., & Wilkins, A.J. (2002). Reading at a distance: Implications for the design of text in children's big books. *British Journal of Educational Psychology, 72*(2), 213-226. https://doi.org/10.1348/000709902158856

Kuster, S.M., van Weerdenburg, M., Gompel, M., & Bosman, A. M. (2018). Dyslexie font does not benefit reading in children with or without dyslexia. *Annals of dyslexia, 68*(1), 25-42. https://link.springer.com/article/10.1007/s11881-017-0154-6

Larson, K., & Carter, M. (2016). Sitka: a collaboration between type design and science. In *Digital Fonts and Reading* (pp. 37-53). https://www.worldscientific.com/doi/abs/10.1142/9789814759540_0003

Legge, G.E. (2006). *Psychophysics of reading in normal and low vision*. CRC Press.

Legge, G.E., & Bigelow, C.A. (2011). Does print size matter for reading? A review of findings from vision science and typography. *Journal of vision, 11*(5), 8-8. https://doi.org/10.1167/11.5.8

Lund, O. (1999). *Knowledge construction in typography: the case of legibility research and the legibility of sans serif typefaces*. Doctoral dissertation, University of Reading.

Lussu, G. (1999). *La lettera uccide: Storie di grafica*. Stampa alternativa & Graffiti.

Mansfield, J.S., Legge, G.E., & Bane, M.C. (1996). Psychophysics of reading. XV: Font effects in normal and low vision. *Investigative Ophthalmology & Visual Science, 37*(8), 1492-1501. https://iovs.arvojournals.org/article.aspx?articleid=2161324

Marinus, E., Mostard, M., Segers, E., Schubert, T.M., Madelaine, A., & Wheldall, K. (2016). A special font for people with dyslexia: Does it work and, if so, why?. *Dyslexia, 22*(3), 233-244. https://doi.org/10.1002/dys.1527

Perondi, L., Gerbino, W., Chia, B., Arista, R., Pignoni, G., & Gaudenzi, G. (2017). Tipografia parametrica e Developmental Dyslexia. *MD Journal, 3*, 88-113. https://dsv.units.it/sites/dsv.units.it/files/ric_grpr/PGBAPG2017.pdf

Price, C.J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage, 62*(2), 816-847. https://doi.org/10.1016/j.neuroimage.2012.04.062

Rand Reading Study Group. (2002). *Reading for understanding: Towards an R&D program in reading comprehension*. http://www.rand.org/content/dam/rand/pubs/monograph_reports/2005/MR1465.pdf

Rastle, K. (2016). Visual word recognition. In *Neurobiology of language* (pp. 255-264). Academic Press. https://doi.org/10.1016/B978-0-12-407794-2.00021-3

Regulation (EU) No 1169/2011. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX-:02011R1169-20180101

Rubin, G.S., Feely, M., Perera, S., Ekstrom, K., & Williamson, E. (2006). The effect of font and line width on reading speed in people with mild to moderate vision loss. *Ophthalmic and Physiological Optics, 26*(6), 545-554. https://doi.org/10.1111/j.1475-1313.2006.00409.x

Sheedy, J.E., Subbaram, M.V., Zimmerman, A.B., & Hayes, J.R. (2005). Text legibility and the letter superiority effect. *Human factors, 47*(4), 797-815. https://doi.org/10.1518/001872005775570998

Shaikh, A.D. (2007). *Psychology of onscreen type: Investigations regarding typeface personality, appropriateness, and impact on document perception*. Doctoral dissertation. https://soar.wichita.edu/handle/10057/1109

Sinico, M., 2015. Tertiary qualities, from Galileo to Gestalt psychology. *History of the Human Sciences*, 28(3), 68-79.

Soleimani, H., & Mohammadi, E. (2012). The Effect of Text Typographical Features on Legibility, Comprehension, and Retrieval of EFL *Learners. English Language Teaching, 5*(8), 207-216. https://eric.ed.gov/?id=EJ1079769

Tarasov, D.A., Sergeev, A.P., & Filimonov, V.V. (2015). Legibility of textbooks: a literature review. *Procedia-Social and Behavioral Sciences, 174*, 1300-1308. https://doi.org/10.1016/j.sbspro.2015.01.751

Tarita-Nistor, L., Lam, D., Brent, M.H., Steinbach, M.J., & González, E.G. (2013). Courier: a better font for reading with age-related macular degeneration. *Canadian Journal of Ophthalmology, 48*(1), 56-62. https://www.sciencedirect.com/science/article/abs/pii/S0008418212004693

Tinker, M.A. (1963). *Legibility of print*. Iowa State University Press.

Velasco, C., & Spence, C. (Eds.). (2018). *Multisensory packaging: Designing new product experiences*. Springer.

Velasco, C., & Spence, C. (2019). The role of typeface in packaging design. In *Multisensory Packaging* (pp. 79-101). Palgrave Macmillan. https://link.springer.com/chapter/10.1007/978-3-319-94977-2_4

Wery, J.J., & Diliberto, J.A. (2017). The effect of a specialized dyslexia font, OpenDyslexic, on reading rate and accuracy. *Annals of dyslexia, 67*(2), 114-127. https://link.springer.com/article/10.1007/s11881-016-0127-1

Xiong, Y.Z., Lorsung, E.A., Mansfield, J.S., Bigelow, C., & Legge, G.E. (2018). Fonts designed for macular degeneration: Impact on reading. *Investigative ophthalmology & visual science, 59*(10), 4182-4189. https://doi.org/10.1167/iovs.18-24334

diid No. 74 — 2021
Doi: 10.30682/diid7421i