Research commissioned by CEPE –
The European Council of the Paint,
Printing Ink and Artists' Colours Industry

Testing legibility at 1.2 and 1.4 mm x-height

Denis G. Pelli, Najib J. Majaj, and Nadine Chahine

GENERAL

RSVP (Rapid Serial Visual Presentation) is a commonly used method to measure reading speed which is a reliable assay of legibility and has been used in reading research for decades (Potter, 1984). The open-source, online-testing web app, EasyEyes.app, developed by Prof. Denis Pelli of NYU allows RSVP testing of online participants. In consultation with Dr. Nadine Chahine and Prof. Najib Majaj, Prof. Denis Pelli of New York University used EasyEyes to test 49 observers recruited online through Prolific.co, with the inclusion criteria of being native and fluent speakers of English. Of the 49 participants, we have complete data for 43. Observers read random five-letter words with an x-height of 1.2 or 1.4 mm at a viewing distance of 30 or 60 cm. We expected that 1.4 mm would be more legible than 1.2 mm at both 30 cm and 60 cm viewing distance.

CONCLUSION

Everyone tested easily read random five-letter words at both 1.2 and 1.4 mm x-height at 30 and 60 cm. Setting text at 1.4mm x-height provided a tiny improvement over 1.2 mm at 60 cm, and none at 30 cm. Increasing x-height from 1.2 to 1.4 mm provided a more than trivial increase in speed to only 1 of 49 observers and only at 60 cm distance. These results were contrary to our expectations and invite further research into the optimal text settings for labels. From these statistically-significant findings, it suggests that there is little appreciable benefit (with regard to legibility) to using fonts with an x-height of 1.4 mm instead of 1.2 mm.

RESULTS

Reading speed is high in all four conditions (1.2 mm and 1.4 mm x-height, and 30 and 60 cm viewing distance) (Table 1). There is a small significant increase in reading speed, from 687 word/min at 1.2 mm to 828 word/min at 1.4 mm at 60 cm (Table 1 and Figure 1). At 30 cm there is no significant effect (Table 1 and Figure 1). The standard error of measurement is consistently about 0.04 (Figure 1), across the four conditions. At 60 cm, there is a significant effect of type size, with the 1.4 mm x-height type being read faster (than 1.2 mm) by a factor of 10^0.08=1.2. At 30 cm there is no significant difference in speed between the two sizes. For each type size there is no significant effect of viewing distance.

Summary statistics across all participants

	condition	word/min	m	se across participants	sd across participants	sd across repetitions	N	parameter
1	RSVP Calibri 1.2 mm @ 60 cm	687	-1.059	0.043	0.301	0.155	49	threshold
2	RSVP Calibri 1.2 mm @ 30 cm	731	-1.086	0.041	0.273	0.078	45	threshold
3	RSVP Calibri 1.4 mm @ 60 cm	828	-1.14	0.034	0.241	0.099	49	threshold
4	RSVP Calibri 1.4 mm @ 30 cm	784	-1.116	0.041	0.278	0.185	45	threshold

Table 1. Summary statistics across all participants.

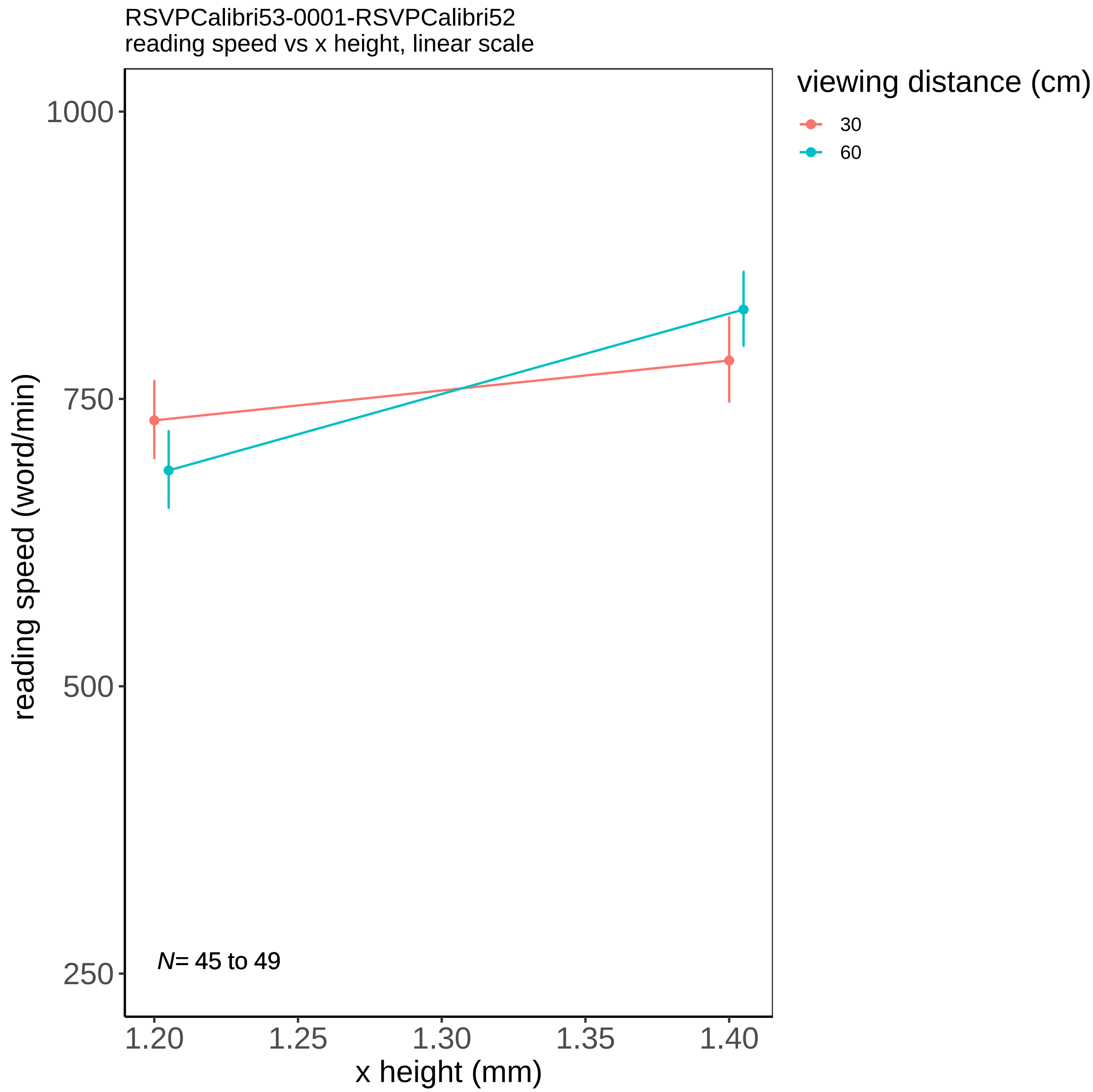

Figure 1. Average reading rate across all observers versus font x-height (mm). There is no effect of size at 30 cm distance (red). There is a significant but negligible result at 60 cm.

It is possible, of course, that a size change that hardly affect most people might still provide an important benefit to a sub population. To examine that we did a scatter diagram, one point per participant, of reading speed at 1.4 mm vs. reading speed at 1.2 mm. At 30 cm, all participants had indistinguishable reading speeds at 1.4 mm vs 1.2 mm (Figure 2). At 60 cm, one person doubled reading speed from roughly 150 to roughly 300 word/min (Figure 3). However, bear in mind that 150 word/min, while slow, would still be plenty to read a paint label in a reasonable amount of time.

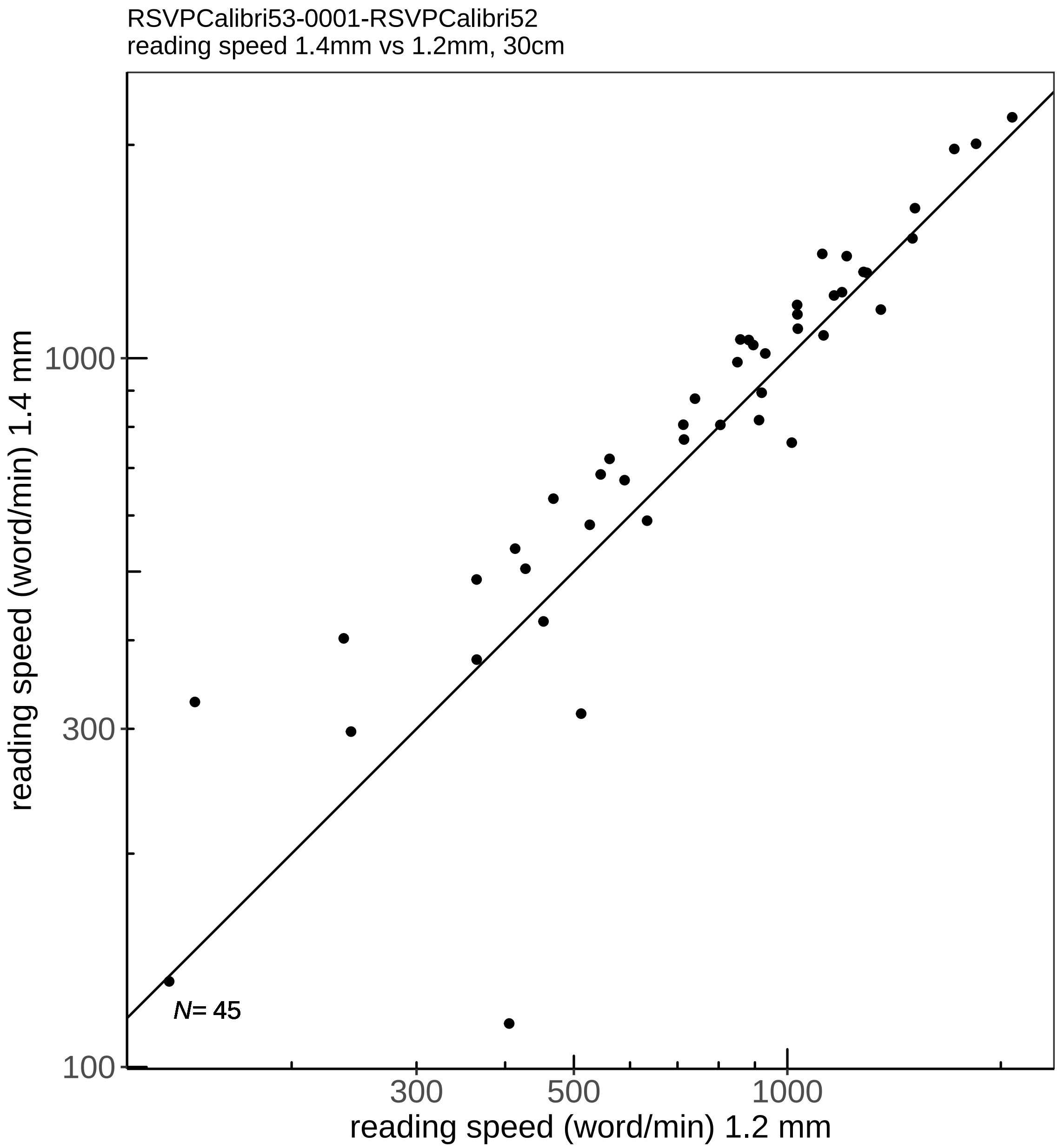

Figure 2. This is a scatter plot, one point per participant, for 30 cm viewing distance, of reading speed with 1.4 vs 1.2 mm. All observers are close to the unity diagonal line, showing that there is no improvement when size changes from 1.2 mm to 1.4 mm

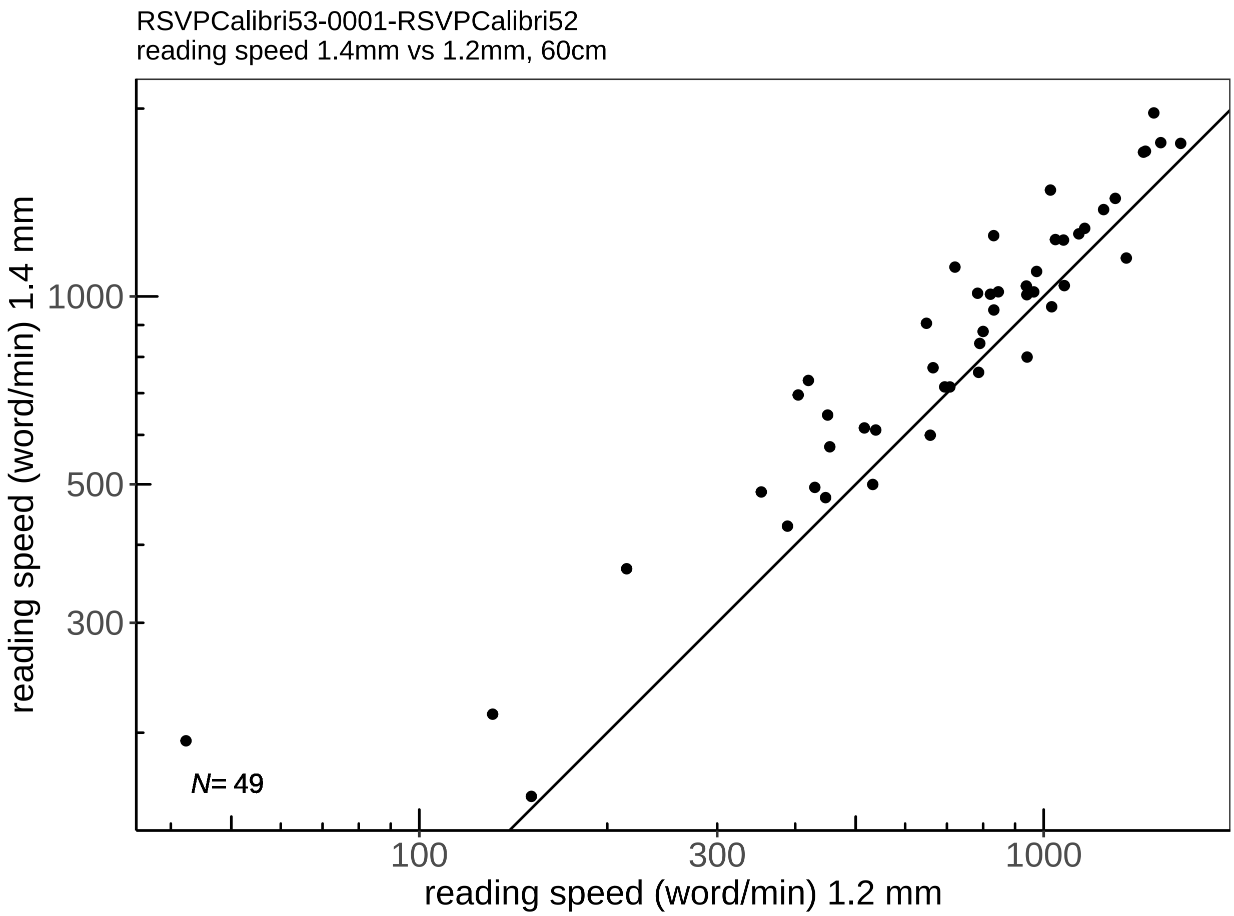

Figure 3. This is a scatter plot, one point per participant, for 60 cm viewing distance, of reading speed with 1.4 vs 1.2 mm. Only one participant showed a large improvement with size (see point in lower left corner). This graph suggests that raising the minimum text size will benefit only 1/50 consumers at 60 cm.

We also tested whether the reading speed difference (between 1.4 and 1.2 mm x-height) was dependent on age and saw no such effect (Figure 4).

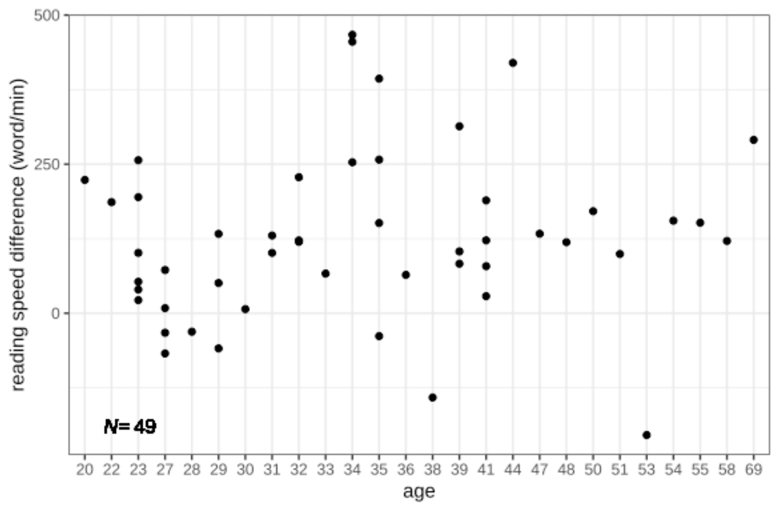

Figure 4. No obvious age-dependence of the reading speed benefit with size (1.4 vs 1.2 mm x-height).

METHODS

All testing was done on the public-domain web app EasyEyes.app created by Prof. Pelli for online vision testing.

Participants. 49 observers were recruited online through Prolific.co, with the inclusion criteria of being native and fluent speakers of English. Prolific requires that participants be at least 18 years old, and their participants are mostly in UK, USA, and Western Europe. Demographic questions asked age (range 20 to 69 years, one person did not state their age) and whether the participant needed glasses/contacts and was using them during testing. Each participant gave informed consent by clicking Yes in response to the project's consent form, which was approved by the New York University Institutional Review Board. Of the 49 participants, we have complete data for 43. Six encountered a fatal error in the software and thus produced incomplete data sets. That is why the value of *N* reported on the graphs varies a bit.

Equipment. As part of the initial compatibility check, EasyEyes required a desktop computer (not tablet or mobile) with at least 6 CPU cores running either the Chrome or Edge browser.

Size and distance. The EasyEyes platform, with help from the participant, measures screen size and viewing distance using the "virtual chinrest" methods of Li et al. (2020). Viewing distance is subsequently monitored continuously using Google FaceMesh processing of the webcam. Each condition had a specified viewing distance of 30 or 60 cm and this was enforced by "nudging". If the observer was more than 20% short or long of the specified distance then the whole screen was occluded by an instruction to move closer or farther as needed.

Lines and words. All text and lines were black on a white background. All test words were displayed in the regular style of the Microsoft Calibri font. The x-height was measured in the participant's browser.

Conditions. Each participant was tested in four conditions, twice. Repeat testing allowed us to assess each participant's SD of test-retest. The four conditions were the combinations of x-height 1.2 or 1.4 mm and viewing distance 30 or 60 cm. Each block tested two conditions randomly interleaved. The 30 cm block interleaved 1.2 and 1.4 mm conditions. Same for the 60 cm block. There were four blocks: 30, 60, 30, 60 cm, with two conditions per block for a total of 8 conditions. The observer did 35 trials in each condition.

Before each trial the observer was asked to fixate the center of a "cross" consisting of a long vertical line and a long horizontal line that would intersect in the middle of the screen except that the lines are suppressed within 1.5 cm of the screen center. Suppressing the fixation lines at screen center prevented any forward masking of the subsequent words. The observer began each trial by pressing the space bar, which immediately provoked the display of three words, one after another, each centered on the screen center.

One trial. Each trial consisted of three random 5-letter words, each flashed for a duration *t*, one after the other, all centered on the screen center. The three words were randomly selected, without replacement from a 380-word corpus consisting of all five-letter words in the Kucera and Francis (1979) corpus with word frequency in the range 50 to 391, so as to exclude very common and rare words. The duration t varies from trial to trial, guided by Quest. After the three words were shown, the observer saw three columns of words, and was asked to click on one word per column, indicating what words had been presented: the first column for the first word and so on. Each column included the word shown and 6 foils selected, without replacement, from the same five-letter word corpus.

Quest. Each word choice was scored right or wrong. Each response was given to Quest along with the duration of the word presentation. Quest implemented a Bayesian procedure to efficiently estimate the duration that would yield 70% correct. For each condition it accumulated the right/wrong responses at each duration. Assuming a Weibull psychometric function and a Gaussian prior probability on log duration it computed the probability density function of the threshold duration at which proportion correct is 70%. Each trial used a duration for each word recommended by Quest, which provided the optimal quantile of the posterior probability density. This optimum maximizes expected information gain of each trial in reducing the entropy of the posterior probability density function. The 35 trials each had three words so each final threshold estimate is based on 105 right/wrong word identifications. The final threshold estimate is the mean of the posterior probability density function. The final threshold estimate is recorded as a log duration. Our plots convert the word duration to words per minute.

FUNDING
This study was funded by CEPE - The European Council of the Paint, Printing Ink and Artists' Colours Industry.

REFERENCES

Francis, W. N., & Kucera, H. (1979). Brown corpus manual. *Letters to the Editor*, *5*(2), 7.

Li, Q., Joo, S. J., Yeatman, J. D., & Reinecke, K. (2020). Controlling for participants' viewing distance in large-scale, psychophysical online experiments using a virtual chinrest. *Scientific reports*, *10*(1), 904.

Potter, M. C. (1984). Rapid serial visual presentation (RSVP): A method for studying language processing. In *New methods in reading comprehension research*. Routledge.

ABOUT THE AUTHORS

Denis G. Pelli
Prof. Denis Pelli is a professor of psychology and neural science at New York University studying object recognition and reading. Pelli studied applied math at Harvard and completed his PhD in physiology at Cambridge. He is known for his contributions to the fields of visual sensitivity, letter identification, object recognition, the Psychtoolbox, equivalent input noise, the Pelli-Zhang video attenuator, and the Pelli-Robson Contrast Sensitivity Chart, which allows for clinical measurement of contrast sensitivity. Current research in Pelli's lab covers object recognition and visual crowding, as well as the experience of beauty. Pelli serves as an associate editor for the Journal of Vision and has published over 50 publications. He has received several awards including the Hopkins Leadership Award from OSA and the Oberdorfer Low Vision Award from the Association for Research in Vision and Ophthalmology. He was a Visiting Fellow Commissioner at Trinity College, Cambridge University 2011 to 2012.
More at https://denispelli.com

Najib J. Majaj
Dr. Najib Majjaj is an Assistant Research Professor at New York University (NYU) and a regular collaborator with Prof. Pelli on various research projects. He has a PhD from NYU and a postdoc from Massachusetts Institute of Technology (MIT). His areas of expertise are deep learning, object recognition, and motion perception with a focus on computational neuroscience. He is also the joint director of The Visual Neuroscience Laboratory which is devoted to the study of the function and development of the mammalian visual system. The laboratory is located in the Center for Neural Science at NYU.

Nadine Chahine
Dr. Nadine Chahine is the CEO at I Love Typography Ltd and principal at ArabicType Ltd. She has an MA in Typeface Design from the University of Reading, a PhD from Leiden University, and an MSt. in International Relations from Cambridge University. Nadine's research focus is on eye movement and legibility studies for the Arabic, Latin, and Chinese scripts. She collaborated with

the MIT Agelab on several legibility studies that have since been published in peer-reviewed journals such as Displays, Ergonomics, and Applied Ergonomics. She was also the lead author of the most recent changes to legibility requirements for text in automotive settings for the ISO 15008 international standard. In 2012 Nadine was selected to Fast Company's 100 Most Creative People in Business and in 2017 to Creative Review's Creative Leaders 50. She has also won multiple design awards including two Awards for Excellence in Type Design from the Type Directors Club in New York.
More at https://ilovetypography.com/research/