Empirical Assessment of the Legibility of the Highway Gothic and Clearview Signage Fonts

Jonathan Dobres, Susan T. Chrysler, Benjamin Wolfe, Nadine Chahine, and Bryan Reimer

Older drivers represent the fastest-growing segment of the driving population. Aging is associated with well-known declines in reaction time and visual processing, and, as such, future roadway infrastructure and related design considerations will need to accommodate this population. One potential area of concern is the legibility of highway signage. FHWA recently revoked an interim approval that allowed optional use of the Clearview typeface in place of the traditional Highway Gothic typeface for signage. The legibility of the two fonts was assessed with color combinations that maximized the contrast (positive or negative) or approximated a color configuration used in highway signage. Psychophysical techniques were used to establish thresholds for the time needed to decide accurately—under glancelike reading conditions—whether a string of letters was a word, as a proxy for legibility. These thresholds were lower for Clearview (indicating superior legibility) than for Highway Gothic across all conditions. Legibility thresholds were lowest for negative-contrast conditions and highest for positive-contrast conditions, with colored highway signs roughly between the two extremes. These thresholds also increased significantly across the age range studied. The method used to investigate the legibility of signage fonts adds methodological diversity to the literature along with evidence supporting the superior legibility of the Clearview font over Highway Gothic. The results do not suggest that the Clearview typeface is the optimal solution for all signage but they do indicate that additional scientific evaluations of signage legibility are warranted in different operating contexts.

A decline in automotive fatalities and major injuries during the past half-century reflects improvements in safety technologies. Little evidence suggests that over that period, human beings somehow have become fundamentally better at the act of driving, which would require unprecedented changes in human vision, motor control, and information processing. Human behavior behind the wheel has changed only as automation (power steering, automatic transmission, and so forth) has eased the difficulties of controlling vehicles and as pervasive and long-ranging educational and enforcement campaigns have induced drivers to wear seatbelts and avoid driving while intoxicated or distracted.

Vehicles also have become safer, through the adoption of passive safety improvements (e.g., crumple zones and composite materials) or the incorporation of new active safety technologies (e.g., electronic stability control and automatic emergency braking). Roadway environment improvements in the areas of pavement and roadside safety hardware include rumble strips and better road surfaces. Traffic engineering improvements to operational elements and traffic control devices also contribute to increased safety. Retroreflective material on traffic signs has steadily increased in brightness since its introduction in the 1930s. Brighter materials improve sign detection and legibility, particularly for older drivers (*1*).

BACKGROUND

Driver Demographics

Highway infrastructure improvements that essentially engineer safety around the driver will be increasingly important in the near future as the world becomes grayer demographically (*2–4*). Older drivers inevitably experience degradations in visual processing, reaction time, and motor control. At the same time, older drivers view the ability to drive as synonymous with independence and a full life and are resistant to forfeiting a driving license, even as they lose the ability to drive safely (*5*). Thus, older drivers would benefit disproportionately from any safety improvement to the vehicle environment that would allow them to continue to drive without sacrificing their safety or that of other drivers. To address the needs of older drivers, FHWA recommends the use of microprismatic sheeting on overhead and ground-mounted guide signs and urges agencies to consider using the Clearview font for positive-contrast legends (*6*).

Signage Typeface Comparisons

Historically, FHWA has mandated the use of a typeface family known as the Standard Highway Alphabet—commonly called Highway Gothic—for road signage (*7*). Developed in the late 1940s and early 1950s, this font family has been in continuous use in the United States for more than 60 years (*8, 9*). An investigation of the legibility of Highway Gothic among older drivers suggested that the simplest way to counteract the effects of the aging eye would be to increase letter size by 20% (*10*). However, implementing this suggestion would be cost prohibitive; large signs would need new structures and possibly

J. Dobres and B. Wolfe, E40-291, and B. Reimer, E40-279, Massachusetts Institute of Technology AgeLab and New England University Transportation Center, 77 Massachusetts Avenue, Cambridge, MA 02139. S. T. Chrysler, Texas A&M Transportation Institute, 3135 TAMU, College Station, TX 77853-3135. N. Chahine, Monotype Imaging, Inc., 600 Unicorn Park Drive, Woburn, MA 01801. Corresponding author: J. Dobres, jdobres@mit.edu.

Transportation Research Record: Journal of the Transportation Research Board, No. 2624, 2017, pp. 1–8.
http://dx.doi.org/10.3141/2624-01

relocation. The Clearview typeface was developed in the early 1990s specifically to improve road sign legibility without increasing overall sign size. On the basis of encouraging preliminary findings that indicated the legibility of Clearview was superior to that of Highway Gothic, FHWA granted interim approval for the use of Clearview on certain types of highway signage in 2004 (*11*). In early 2016, FHWA discontinued the approval, citing the ambiguity of subsequent research results (*12*).

The interim approval allowed Clearview as an alternate typeface for positive-contrast signs (white letters on a dark background). In its notice of termination, FHWA cites studies that tested Clearview in forms that never were allowed in the interim approval [e.g., condensed letterform (*13*) and negative contrast (*14*)]. However, the notice of discontinuation overlooks most research on Clearview. Even though some studies have found Clearview not to hold a legibility advantage (*13–15*), many others have reported significant evidence that Clearview is more legible (*16–23*). Even Miles et al., who dismiss Clearview in the summary of their evaluation, found results that could be described as mixed at worst: Clearview improved legibility distances for older drivers during the day but not at night (*15*). The FHWA notice also states that most of the apparent legibility benefit was attributable to the use of newer, brighter sheeting materials. Even though many studies on sheeting type have indicated that it contributes more to legibility improvements than font, statistical analyses of several studies indicate that Clearview is a significant contributor to improvements in legibility, regardless of the sheeting used (*13, 16, 18, 20, 22*).

Studies on Clearview have shown legibility improvements in the range of 2% to 30% (or decrements of as much as 7% among studies that failed to find Clearview superior). This range suggests that the effect of the typeface is subtle compared with other factors. It is especially relevant given that all studies of Clearview to date have used similar investigatory methods and therefore suffer from the same set of limitations. In such studies, participants drive along a test track and verbally report the content of signs posted in various positions. The distance at which the driver correctly reports the content of a given sign is taken as the measure of its legibility.

The real-world validity of such assessments has obvious advantages, but many limitations and uncontrolled variables apparent in this general design can complicate interpretation of results. Several studies distinguish day and night driving (*14–17, 21*), but only one reports an attempt to account for whether the day was sunny or overcast (*22*), which can reduce the amount of available illumination by as much as half. Results also may be affected by the particularities of the test track and surrounding area, the type and size of vehicle driven, whether the research participant is driving or reading signs as a passenger, and how fast the vehicle is traveling—which ranges from 8 km/h (5 mph) to 56 km/h (35 mph) in these studies (*14, 15*). All but one of the aforementioned studies relies on participants to report responses verbally, which introduces unavoidable and unpredictable delays between stages, including the driver's accurate perception of the sign, initiation of a verbal response, the experimenter's recognition of the response, and recording of the legibility distance (*17*). These delays introduce an uncontrolled source of noise in the data, especially with the vehicle traveling at a high speed relative to human-scale action. In addition, the need to fabricate physical signs for testing means that each research participant will see only a handful of trials in each tested condition, which by definition reduces the statistical power of the experiment (*24*). Whereas controlled field tests like these are the norm in traffic engineering assessments of legibility, psychological investigations of text legibility have used more tightly controlled, laboratory-based methods of assessment.

What Makes Text Legible?

Slattery and Rayner define legibility as the ease with which a reader can accurately perceive and encode text (*25*). Legibility is a product of intrinsic factors (which include character shape, width, and weight) and extrinsic factors (which include overall size, illumination, contrast, color, and polarity—that is, whether the text is dark against a lighter background or vice versa). Most studies cited earlier report that the legibility of Clearview is superior to that of Highway Gothic over a range of extrinsic factors. To understand why such results would be expected, one first must understand which intrinsic factors make one typeface more or less legible than another.

To the casual observer, many typefaces may appear similar or nearly indistinguishable. However, the appearance of a well-designed typeface is governed by many well-known parameters, a few of which are described in this paper; others are described elsewhere (*26*). Figure 1*a* shows the lowercase "c" character in 13 sans serif typefaces, all rendered at a nominal font height of 100 points. Among these samples, height variations arise from the different ratios of x-height (i.e., lowercase letter height) intrinsic to each font's design. The line thickness (or stroke weight) varies considerably and may increase or decrease over a contour to give an appearance of flaring. The character openness varies, the extremes of which are illustrated in the Eurostile and Frutiger typefaces (far left and far right, respectively).

Figure 1*b* illustrates other important variables in typeface design. Even though the two typefaces are scaled to identical capital letter heights in the figure, Clearview appears larger than Highway Gothic because its x-height is much higher and its ascender is taller (notable in lowercase "i" and "h," which exceed capital letter height). Clearview and Highway Gothic also have different stroke terminations, seen most prominently in the tail of the lowercase "g." Finally, Clearview has much larger counters (i.e., the partially or completely enclosed spaces in certain letterforms), as illustrated by the lowercase "a," "e," and "g."

These differences are not trivial. Even though one might be tempted to consider them as merely aesthetic, stroke terminations reported are a key determinant of legibility (*27, 28*). The perceived text size is a crucial mediator of legibility, as one might expect. Therefore, increasing the x-height is one way to improve legibility (*29–32*). The use of capital letters does not achieve the same advantage. Even though their larger size makes them more legible, capital letters have been shown to be more easily confused (*33–35*). In addition, typefaces with more varied shapes and open spacing have been shown to be more legible than typefaces with more uniform shapes and closed spacing (*26, 32, 36*). This finding is especially relevant to typography meant for use on road signage. Letters illuminated by headlights at night exhibit halation (i.e., a substantial glowing effect that blurs the boundaries of the letters). Typefaces with more open letterforms, less uniform shapes, and larger x-heights resist visual degradation from the effect of halation and therefore can be read more easily at night.

Empirical Assessments of Legibility

Laboratory-based research on text legibility has been of interest for nearly as long as the discipline of psychological science has existed,

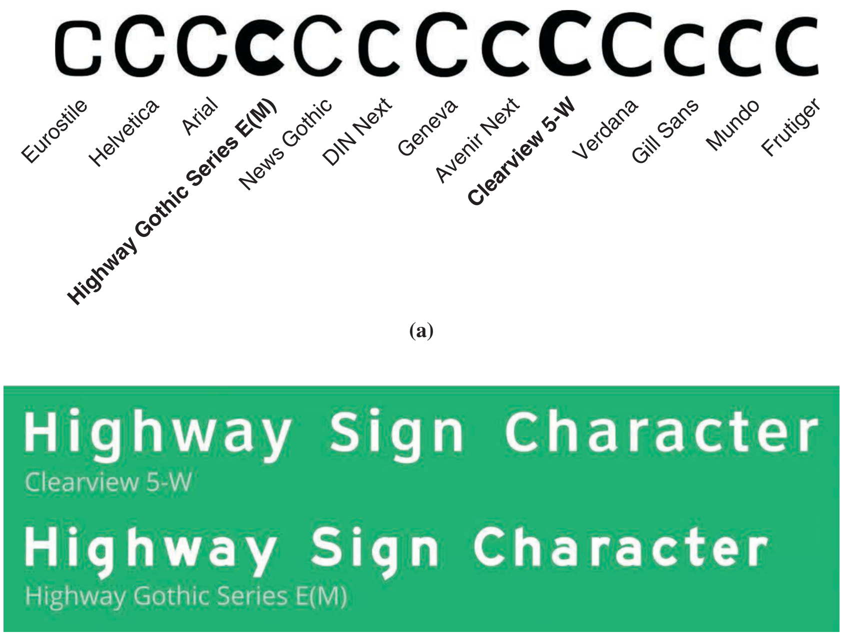

FIGURE 1 Samples of (*a*) lowercase "c" in 13 fonts, arranged in order of openness from left (less open) to right (more open), adapted from Reimer et al. (*26*), and (*b*) Clearview 5-W and Highway Gothic Series E-Modified typefaces as they might appear on typical highway signage (color approximate) [E(M) = E-Modified].

and such methods have been used to examine various factors that influence legibility (*37*). [Extensive reviews are available elsewhere (*26, 32*).] As in-vehicle human–machine interfaces have increased in complexity in recent years, so has the primacy of digital text in these interfaces and interest in legibility. Studies with an automotive focus have examined legibility factors such as type size and text contrast and brightness (*30, 31*).

Studies of text legibility that examine the intrinsic characteristics of typeface design are rare in automotive and other contexts. A study conducted in a full-cab driving simulator, in which human–machine interface menus were set in one of two possible typefaces, found that typeface choice was correlated with significant differences in driver behavior, as measured by task completion time, errors made with the device, and time spent glancing at the device screen (*26*). Follow-up research on these typefaces that used a desktop-based psychophysical method (lexical decision) successfully emulated the glancelike reading behavior common while driving and confirmed the results of the previous study (*32*). Because typographic configurations are presented in a controlled setting, these methods offer a more direct measure of text legibility, free of environmental or situational confounds. The studies suggest that humanist-style typefaces—the design aesthetics of which favor open shapes, loose spacing, varied letterforms, and other features similar to Clearview—improve the legibility of written materials.

In this study, the legibility of Clearview and Highway Gothic typefaces was investigated with a lexical decision task (*38*). The performance measure in this task is the on-screen display time required for a reader to decide accurately whether a presented stimulus is a word (i.e., the lexical decision). In an adaptive staircase procedure, the display time of the letter string is manipulated until a targeted accuracy is attained. The single-word display time required to produce approximately 80% accuracy in the lexical decision task is called the legibility threshold. Typefaces that are more legible should have lower legibility thresholds, meaning that a reader requires less time to read the words at the targeted accuracy level.

The study objective was to establish legibility thresholds for Clearview and Highway Gothic typefaces under conditions that either mimic a common signage configuration or are designed to gauge theoretical legibility under maximal contrast. The authors hypothesized that (*a*) the legibility of Clearview would be superior to that of Highway Gothic across all conditions, (*b*) legibility thresholds would increase with age, and (*c*) the effect of aging on text legibility would be more pronounced for Highway Gothic than for Clearview.

METHODS

Participants

Participants were recruited from the Boston and Cambridge, Massachusetts, area for the study. Of 34 total participants, the results of four were excluded from the final sample: three had outlier legibility thresholds greater than 300 ms, and one exceeded the desired gender balance. The remaining 30 participants were aged between 36 and 74 years old. All participants reviewed and signed an approved informed consent form.

All participants were required to be between 35 and 75 years old (pilot research had suggested that differences in lexical decision performance attributable to typeface design become apparent starting in the 30s), in reasonably good health for their age (as self-reported), and native English speakers. The vision of all participants was normal or corrected-to-normal (with glasses or contact lenses) and was tested on site with the FAA's Form 8500-1 for near acuity and a Snellen eye chart for far acuity. All participants had a Snellen acuity of at least

20/40 at a standard 20-ft (6-m) distance. The mean age of the sample was 55.4 years, and the mean age of the male and female subsamples (15 participants each) was identical [$t(28) = 0.0$, $p = 1.0$].

Apparatus, Task, and Stimuli

The experiment was conducted in a quiet, dimly lit room. Participants were positioned such that their eyes were approximately 70 cm (27.5 in.) from the display (27 in. diagonal, 2,560 pixels × 1,440 pixels resolution, 60-Hz refresh rate). Head restraints were not used, allowing a degree of positional variability likely in driving scenarios. Participants were positioned with their eyes 27 in. from the display and permitted to use whatever optical correction felt comfortable at that distance.

Participants were asked to perform a one-interval forced-choice lexical decision task that entailed making a simple yes-or-no decision about whether a string of text presented briefly formed a common English word or a pseudoword (i.e., a letter string that is pronounceable but not a word) (*32, 36, 38*). Figure 2*a* is a schematic of one lexical decision trial, which begins with the display of a fixation rectangle for 1,000 ms; all subsequent stimuli are presented at the center of this rectangle. The 700- × 400-pixel rectangle centered at the midpoint of the screen is filled with the chosen background color; the surrounding screen is filled with pure black to minimize illumination differences between conditions. Next, a masking stimulus composed of randomized characters (chosen from among the possible characters XxYy=) is displayed for 200 ms, followed by the target stimulus for a variable duration, as determined by an adaptive staircase procedure (described later). Another 200-ms mask then is displayed, followed by a response screen. Participants are given up to 5,000 ms to make the lexical decision (i.e., word or pseudoword) by pressing one of two keys on the computer's numeric keypad. Participants receive feedback on response accuracy only during the practice section (described later).

The experiment began with a series of 10 practice trials in which stimulus duration was set to 1,000 ms. Each participant was permitted to move on to the main experiment after five consecutive correct answers; those who completed 10 trials without five consecutive correct responses were allowed to repeat the practice block.

Stimuli were displayed in all lowercase lettering. Fonts were scaled to a 4-mm-high capital letter, consistent with previous studies (*26, 32*). Georgia, a serif typeface that looks substantially different from the two typefaces of interest, was used at twice the letter height to display practice trial stimuli and all prompt text.

Words and pseudowords were generated from an online orthographic database (*39*). All words and pseudowords were six characters long, and search parameters were constrained to provide lists of reasonably common English words. [Details have been published elsewhere (*32*).]

Adaptive Staircase Procedure

An adaptive staircase procedure was used to control the difficulty of the lexical decision task by adjusting the stimulus duration (on-screen display time for words and pseudowords) according to the response accuracy of the participant (*32, 36, 40, 41*). This experiment followed a three-down, one-up rule: stimulus duration was reduced after three consecutive correct responses; display duration was increased after one incorrect response. According to this rule, the staircase converges on a stimulus duration that corresponds to approximately 80% decision accuracy. The 80% threshold was chosen from among several possible options because pilot testing had indicated that research participants found the task frustrating with lower accuracy points. (Conversely, higher accuracy points would have made the task too easy.) Therefore, stimulus duration is the primary dependent measure of legibility; when the display is more legible, a participant should attain target accuracy with a shorter stimulus duration. The chosen accuracy point is low enough to prevent the task from being too easy but high enough to avoid excessive frustration. Because all test conditions are calibrated to the same theoretical accuracy point, typefaces with better legibility should have lower threshold stimulus durations.

Each typographic condition was assessed in a block of 100 consecutive trials. Under each condition, stimulus duration was predetermined for the first few trials before beginning staircase control. Stimulus duration began at 800 ms. After three trials, it was reduced to 600 ms, then 400 ms, and finally 200 ms. This controlled descent allowed participants to adapt to the expected task difficulty level. Then, staircase control was activated, and stimulus duration changed automatically according to participant responses. Stimulus duration was bounded between 16.7 ms (the lower limit of the screen) and 1,000 ms. In the

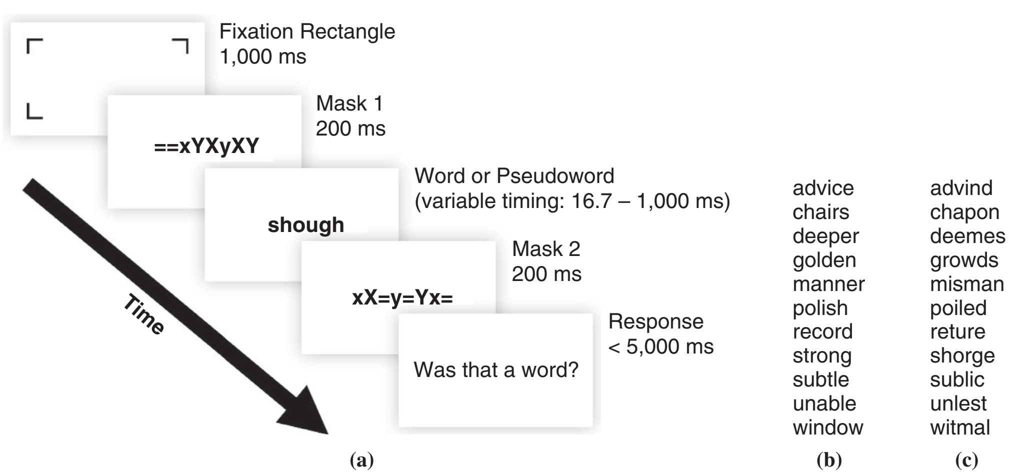

FIGURE 2 Lexical decision trial (*a*) structure, (*b*) example words, and (*c*) example nonwords. (Examples are illustrated in Open Sans typeface, which was not used in the experiment.)

experimental sample, actual stimulus durations were between 16.7 and 450 ms (median of 100 ms).

Conditions Assessed

Legibility thresholds were assessed for text set in two typeface families and in three color combinations for a total of six conditions per participant. Color combinations included maximal negative contrast (black text on white background), maximal positive contrast (white text on black background), and a color combination similar to one used for highway signage (white text on green background). The exact shade of green was determined during study design by matching the on-screen color to a Pantone 342C color chip illuminated by approximately 500 lux; 14 individuals created subjective color matches to the chip, and the resulting values were averaged. Screen renderings are presented in Figure 3, in which the green shade used for the highway sign condition appears darker in print than under the screen illumination used in the experiment.

Clearview and Highway Gothic type families were assessed. Both families include variants meant for use on different signage types. Highway Gothic Series E-Modified was chosen as representative of Highway Gothic because it is used pervasively on white-on-green highway guide signage. Its Clearview counterparts, Clearview 5-W and 5-B, have slightly different stroke weights and are intended for use when sign text is white and black, respectively. Therefore, Highway Gothic E-Modified was used in all three Highway Gothic conditions, and Clearview 5-W or 5-B was used for conditions with white or black text, respectively. Condition order was randomized by participant to counteract the possible effects of habituation and learning.

Data Reduction and Analysis

Three dependent measures were collected for each trial: stimulus duration (i.e., word or pseudoword display time), response accuracy, and response time. Stimulus duration thresholds (legibility thresholds) were calculated as the median stimulus duration of each condition's final 20 trials. Response accuracy for the final 20 trials of each condition were averaged to assess the stabilization of the staircase procedure. Mixed-effect linear modeling techniques were used to analyze and visualize data in R (*42*). The response time data are not presented because the study focus was threshold stimulus display times required to achieve the target accuracy.

RESULTS

Response Accuracy

As expected when adaptive staircase methods are used, accuracy did not differ significantly between conditions [$F(5, 145) = 0.37$, $p = .868$) and accuracy was not significantly different from the theoretical calibration point of 80% (all $p > .253$, one-sample *t*-tests) under any of the six conditions.

Legibility Thresholds

Figure 4*a* shows the mean legibility thresholds assessed for each condition studied. Lower values indicate that the lexical decision was made with the target level of accuracy in less time and are taken as a proxy for greater legibility. A linear mixed-effect model was constructed that included participant age, sign color (positive contrast, negative contrast, or guide sign coloring), and typeface (Clearview or Highway Gothic) as fixed effects and participant as a random effect.

Results indicate that sign coloring was correlated with significant differences in legibility thresholds under laboratory conditions [$F(2, 130) = 7.13$, $p = .001$]. Posthoc testing indicated that negative contrast was significantly different from positive contrast and guide sign coloring [$t(29) = 4.33$, $p < .001$ and $t(29) = 2.50$, $p = .018$, respectively] and that guide sign coloring was not significantly different from positive contrast [$t(29) = 1.50$, $p = .144$].

Typeface also had a significant effect on legibility thresholds [$F(1, 130) = 8.85$, $p = .003$]: thresholds were lower for text set in Clearview (i.e., indicating better legibility) across all conditions. Typeface and sign color did not interact significantly [$F(2, 130) = 0.17$, $p = .842$], suggesting that the effect of typeface was similar across color combinations.

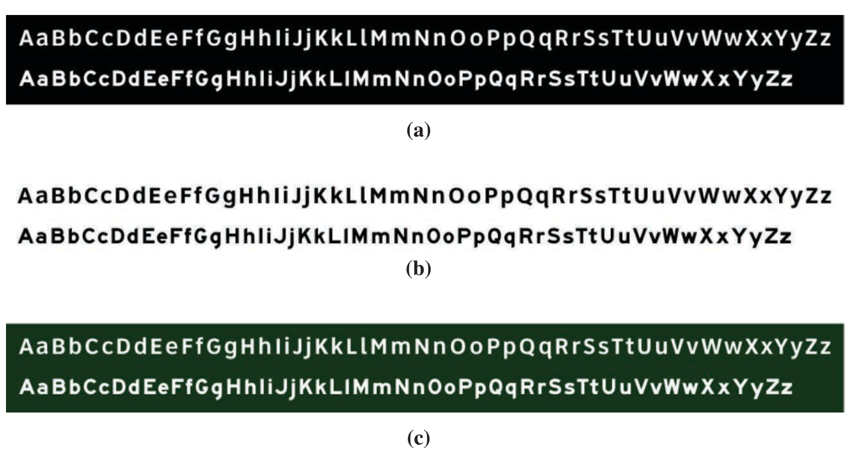

FIGURE 3 Text samples, as rendered at the size and on the monitor used (*a*) at maximum negative contrast, (*b*) at maximum positive contrast, and (*c*) with "highway sign" coloring (Clearview 5-W or 5-B in odd rows, Highway Gothic Series E-Modified in even rows).

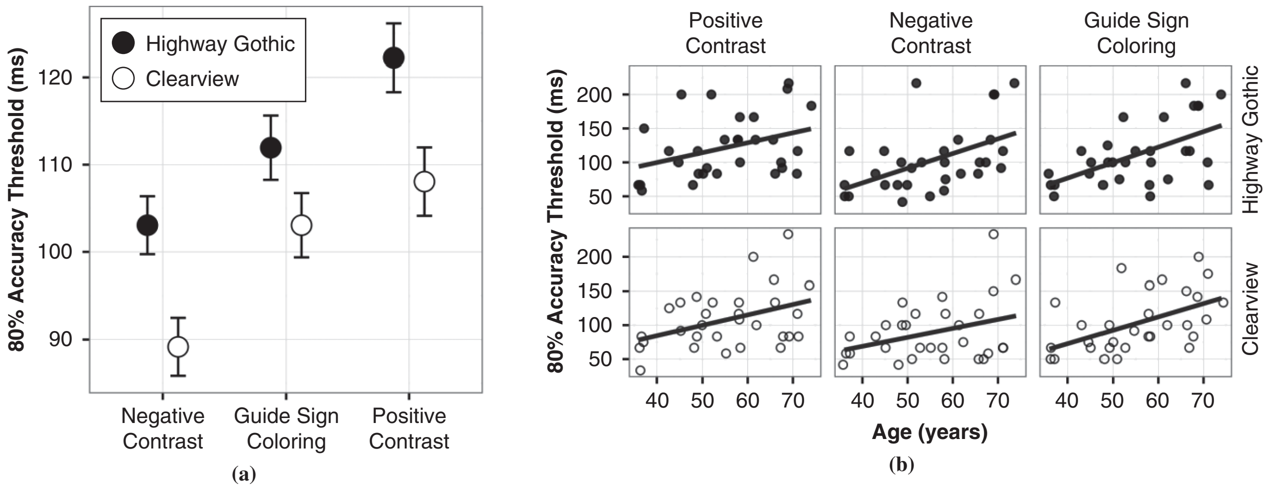

FIGURE 4 Legibility thresholds corresponding to 80% word recognition accuracy under all conditions: (*a*) means (lower numbers indicate better legibility; error bars represent ±1 mean-adjusted standard error) and (*b*) for individual participants, visualized as a function of participant age under all conditions tested (black lines indicate simple linear regressions through the data points; points have been jittered slightly to minimize overlap).

The analysis results in Figure 4*b* indicate that legibility thresholds increased significantly with age [$F(1, 26) = 12.01$, $p = .002$], consistent with well-known age-related degradations in visual processing. Correlation was not observed between age and typeface [$F(1, 130) = 0.98$, $p = .325$], sign coloring [$F(2, 130) = 1.01$, $p = .369$], or the combination of all three factors [$F(2, 130) = 0.53$, $p = .592$], which suggests that age-related increases in reading time thresholds were consistent across the study conditions. The linear model shows that in these data, legibility thresholds increased 93%, from an average of 72 ms at age 36 to 139 ms at age 74.

DISCUSSION OF RESULTS

The legibility of commonly used variants of the Clearview and Highway Gothic typeface families was assessed under color combinations that maximized contrast (positive or negative) or approximated a color configuration commonly used in highway signage. Psychophysical techniques were used in the evaluation to establish legibility thresholds, measured as the amount of display time needed to read a single word with approximately 80% accuracy under glancelike reading conditions. Reading time thresholds were lower for Clearview (indicating better legibility) than Highway Gothic across all conditions tested. Thresholds were lowest for negative contrast and highest for positive contrast, with the approximated highway sign coloring demonstrating thresholds between the two extremes. Legibility thresholds also increased significantly across the age range studied. However, in contrast to one author's hypothesis, the rate of increase in legibility threshold across the age range did not differ significantly by typographic condition.

In agreement with most test track studies, results of the study indicate that the legibility of Clearview is superior to that of Highway Gothic. On average, the legibility thresholds of Highway Gothic are 12.3% higher than those of Clearview, indicating that Highway Gothic required more time to read accurately. The legibility of Clearview is superior, regardless of color combination, suggesting that its legibility arises from characteristics intrinsic to its design rather than extrinsic factors (e.g., type size and text contrast and brightness) (*43*). This finding is in agreement with previous results indicating that typefaces with open letterforms, varying shapes, and generous x-height ratios—such as humanist-style typefaces, evaluated in earlier work—are more legible than other typefaces (*26, 32*). Indeed, statistical comparisons with earlier work indicate that stimulus duration thresholds for Clearview and Frutiger (a humanist typeface), assessed under similar size and color conditions, are not significantly different. Had they been scaled to equalize x-heights instead of capital heights, Highway Gothic and Clearview probably would have exhibited similar legibility thresholds because this scaling would have somewhat enlarged the typesetting of Highway Gothic. However, scaling by capital height is specified for signage layout guidelines in the *Manual on Uniform Traffic Control Devices* (*7*) and in guidelines for the typography of in-vehicle devices (ISO 15008:2009, Road Vehicles—Ergonomic Aspects of Transport Information and Control Systems—Specifications and Test Procedures for In-Vehicle Visual Presentation). Reimer et al. illustrate the effect of x-height versus capital height on the scaling of typefaces (*26*).

The data also support the idea that the legibility of negative-contrast color combinations is superior to that of positive-contrast conditions. However, the authors do not believe it to be a byproduct of typographic design. Rather, research in this area suggests that the effect has more to do with the background colors in each condition and the amount of illumination emitted into the observer's pupil from each. Research has shown that legibility is enhanced under conditions of increased illumination, regardless of the colors used, most probably because the contraction of the pupil reduces optical distortion as light enters the eye (*44–46*). Even though the legibility of negative-contrast conditions appears to be superior in the lab, such an advantage must be balanced with the practical disadvantage of halation (i.e., overglow or blurring) encountered on real-world road signage during night driving. Pervasive use of negative-contrast signage would result in extreme halation and probably would outweigh any advantage from the increased illumination.

Legibility thresholds increased dramatically (by an estimated 93%) across the age range studied. This finding is particularly relevant in light of Clearview's origins as a typeface designed in part to make road signage more legible for older drivers. In this study, legibility thresholds increased with age at the same rate, independent of the typeface or color condition. Because the reading time thresholds of Clearview are lower than those of Highway Gothic across the age range studied, Clearview attains the intended goal of increasing legibility for older drivers in a way that benefits all drivers. An average increase of 67 ms may seem small, but this research was designed to emulate reading at a glance, and drivers will glance three or four times per second as they scan the road environment, including signage. If U.S. roadways are to accommodate an increasing number of older drivers, then the signs that make those roadways navigable must considered (*4*).

As indicated in the earlier discussion of the observed effects of contrast, this laboratory-based method differs from on-road research methods in some key ways. Even though laboratory methods can emulate glancelike behavior to some extent, the legibility thresholds measured in this study are considerably lower than glance dwell times observed in typical test track studies and in on-road research. In this sense, the stimulus duration thresholds are more akin to information processing thresholds and more directly related to the pure legibility of the stimuli. In contrast, on-road glance dwell times are influenced by a wider environment of factors and are expected to be longer. Similarly, the halation produced on an electronic screen in this study is negligible compared with what would be encountered during typical night driving, and the authors speculate that the method underestimates the effect of halation, as mentioned earlier.

In another departure from real-world use cases, Highway Gothic E-Modified was used across all conditions, including negative contrast, which ordinarily calls for Series D. Series D has marked design differences from Series E (thinner strokes and narrower, more condensed letterforms). In this sense, the authors believe that Series D would have acted as a third typeface, unbalancing the study design and greatly weakening statistical analyses. The authors conjecture that the more condensed letterforms of Series D would have increased the visual similarity of the letters, decreasing legibility compared with the other typefaces.

Even with these limitations in mind, the authors confidently add the present results to the long series of studies that show Clearview offers superior legibility to the Highway Gothic series. Clearview outperformed Highway Gothic in laboratory research designed to assess basic legibility differences across a range of intrinsic and extrinsic factors, adding converging evidence of its superior legibility.

ACKNOWLEDGMENTS

Support for this work was provided in part by the New England University Transportation Center at the Massachusetts Institute of Technology (U.S. DOT, Region 1), the Southwest Region University Transportation Center, and the Toyota Class Action Settlement Safety Research and Education Program. The authors thank Donald Meeker and James Montalbano for providing access to the Clearview typeface family as well as providing background on the origins and design of the typeface.

REFERENCES

1. Chrysler, S., D. Tranchida, S. Stackhouse, and E. Arthur. Improving Street Name Sign Legibility for Older Drivers. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting,* Vol. 45, No. 23, 2001, pp. 1597–1601. http://dx.doi.org/10.1177/154193120104502308.
2. Roberts, L. 9 Billion? *Science,* Vol. 333, No. 6042, 2011, pp. 540–543. http://dx.doi.org/10.1126/science.333.6042.540.
3. Sade, R. M. The Graying of America: Challenges and Controversies. *The Journal of Law, Medicine, and Ethics,* Vol. 40, No. 1, 2012, pp. 6–9.
4. Stock, K. *How Old People Became the Future of the U.S. Auto Industry.* Bloomberg.com, Aug. 11, 2015. http://www.bloomberg.com/news/articles/2015-08-11/how-old-people-became-the-future-of-the-u-s-auto-industry. Accessed July 20, 2016.
5. Persson, D. The Elderly Driver: Deciding When to Stop. *Gerontologist,* Vol. 33, No. 1, 1993, pp. 88–91. http://dx.doi.org/10.1093/geront/33.1.88.
6. *Handbook for Designing Roadways for the Aging Population.* FHWA, U.S. Department of Transportation, 2014.
7. *Manual on Uniform Traffic Control Devices.* FHWA, U.S. Department of Transportation, 2009. http://mutcd.fhwa.dot.gov/.
8. Loutzenheiser, D. W. Design of Signs for the Pentagon Road Network. *Highway Research Record,* Vol. 23, 1944, pp. 206–235.
9. Forbes, T. W., K. Moscowitz, G. Morgan, and D. W. Loutzenheiser. A Comparison of Lower Case and Capital Letters for Highway Signs. *Highway Research Record 30,* 1951, pp. 355–373.
10. Mace, D. J., P. M. Garvey, and R. F. Heckard. *Relative Visibility of Increased Legend Size Versus Brighter Materials for Traffic Signs.* FHWA-RD-94-035. FHWA, U.S. Department of Transportation, 1994, pp. 1–57.
11. McElroy, R. S. *Manual on Uniform Traffic Control Devices: Interim Approval for Use of Clearview Font for Positive Contrast Legends on Guide Signs.* FHWA, U.S. Department of Transportation, Sept. 2, 2004. http://mutcd.fhwa.dot.gov/res-ia_clearview_font.htm. Accessed July 14, 2016.
12. National Standards for Traffic Control Devices; the Manual on Uniform Traffic Control Devices for Streets and Highways; Notice of Termination of Interim Approval IA–5. *Federal Register,* Vol. 81, No. 15, Jan. 25, 2016, pp. 4083–4084.
13. Chrysler, S. T., P. J. Carlson, and H. G. Hawkins, Jr. *Nighttime Legibility of Ground-Mounted Traffic Signs as a Function of Font, Color, and Retroreflective Sheeting Type.* FHWA/TX-03/1796-2. Texas Transportation Institute, College Station, 2002, pp. 1–76.
14. Holick, A. J., S. T. Chrysler, E. S. Park, and P. J. Carlson. *Evaluation of the Clearview Font for Negative Contrast Traffic Signs.* FHWA/TX-06/0-4984-1. Texas Transportation Institute, College Station, 2006, pp. 1–130.
15. Miles, J. D., B. Kotwal, S. Hammond, and F. Ye. *Evaluation of Guide Sign Fonts. Publication MN/RC 2014-11.* Texas A&M Transportation Institute, College Station, 2014, pp. 1–61.
16. Garvey, P., M. Pietrucha, and D. Meeker. Effects of Font and Capitalization on Legibility of Guide Signs. *Transportation Research Record,* No. 1605, 1997, pp. 73–79. http://dx.doi.org/10.3141/1605-09.
17. Hawkins, H., Jr., D. Picha, and M. Wooldridge. Performance Comparison of Three Freeway Guide Sign Alphabets. *Transportation Research Record: Journal of the Transportation Research Board,* No. 1692, 1999, pp. 9–16. http://dx.doi.org/10.3141/1692-02.
18. Carlson, P., and G. Brinkmeyer. Evaluation of Clearview on Freeway Guide Signs with Microprismatic Sheeting. *Transportation Research Record: Journal of the Transportation Research Board,* No. 1801, 2002, pp. 27–38. http://dx.doi.org/10.3141/1801-04.
19. Holick, A. J., and P. J. Carlson. *Nighttime Guide Sign Legibility for Microprismatic Clearview Legend on High Intensity Background.* FHWA/TX-04/0-1796-4. Texas Transportation Institute, College Station, 2003, pp. 1–116.
20. Carlson, P. J., and A. Holick. Maximizing Legibility of Unlit Freeway Guide Signs with Clearview Font and Combinations of Retroreflective Sheeting Materials. *Transportation Research Record: Journal of the Transportation Research Board,* No. 1918, 2005, pp. 26–34. http://dx.doi.org/10.3141/1918-04.
21. Garvey, P. M., A. Z. Zineddin, and M. T. Pietrucha. Letter Legibility for Signs and Other Large-Format Applications. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting,* Vol. 45, No. 18, 2001, pp. 1443–1447.
22. Gowda, R. N. *Evaluation of the Effect of Clearview Font and Retro-Reflective Sheeting Materials on Legibility Distance.* MS thesis. Kansas State University, Manhattan, 2010.
23. Guerrier, J. H., and S.-H. Fu. *Elder Roadway User Program Test Sections and Effectiveness Study.* FDOT 669535. University of Miami School of Medicine, Tallahassee, Fla., 2002, pp. 1–204.
24. Cohen, J. A Power Primer. *Psychological Bulletin,* Vol. 112, No. 1, 1992, pp. 155–159. http://dx.doi.org/10.1037/0033-2909.112.1.155.
25. Slattery, T. J., and K. Rayner. Influence of Text Legibility on Eye Movements During Reading. *Applied Cognitive Psychology,* Vol. 24, No. 8, 2010, pp. 1129–1148. http://dx.doi.org/10.1002/acp.1623.

26. Reimer, B., B. Mehler, J. Dobres, J.F. Coughlin, S. Matteson, D. Gould, N. Chahine, and V. Levantovsky. Assessing the Impact of Typeface Design in a Text-Rich Automotive User Interface. *Ergonomics,* Vol. 57, No. 11, 2014, pp. 1643–1658. http://dx.doi.org/10.1080/00140139.2014.940000.
27. Attneave, F., and M.D. Arnoult. The Quantitative Study of Shape and Pattern Perception. *Psychological Bulletin,* Vol. 53, No. 6, 1956, pp. 452–471. http://dx.doi.org/10.1037/h0044049.
28. Fiset, D., C. Blais, C. Ethier-Majcher, M. Arguin, D. Bub, and F. Gosselin. Features for Identification of Uppercase and Lowercase Letters. *Psychological Science,* Vol. 19, No. 11, 2008, pp. 1161–1168. http://dx.doi.org/10.1111/j.1467-9280.2008.02218.x.
29. Legge, G.E., and C.A. Bigelow. Does Print Size Matter for Reading? A Review of Findings from Vision Science and Typography. *Journal of Vision,* Vol. 11, No. 5, 2011.
30. Fujikake, K., S. Hasegawa, M. Omori, H. Takada, and M. Miyao. Readability of Character Size for Car Navigation Systems. In *Human Interface and the Management of Information: Interacting in Information Environments.* Springer, Berlin, 2007, pp. 503–509.
31. O'Day, S., and L. Tijerina. Legibility: Back to the Basics. *SAE International Journal of Passenger Cars: Mechanical Systems,* Vol. 4, No. 1, 2011, pp. 591–604.
32. Dobres, J., N. Chahine, B. Reimer, D. Gould, B. Mehler, and J.F. Coughlin. Utilising Psychophysical Techniques to Investigate the Effects of Age, Typeface Design, Size, and Display Polarity on Glance Legibility. *Ergonomics,* Vol. 59, No. 10, 2016, pp. 1377–1391.
33. Bouma, H. Interaction Effects in Parafoveal Letter Recognition. *Nature,* Vol. 226, No. 5241, 1970, pp. 177–178. http://dx.doi.org/10.1038/226177a0.
34. Bouma, H. Visual Interference in the Parafoveal Recognition of Initial and Final Letters of Words. *Vision Research,* Vol. 13, No. 4, 1973, pp. 767–782. http://dx.doi.org/10.1016/0042-6989(73)90041-2.
35. Pelli, D.G., K.A. Tillman, J. Freeman, M. Su, T.D. Berger, and N.J. Majaj. Crowding and Eccentricity Determine Reading Rate. *Journal of Vision,* Vol. 7, No. 2, 2007, pp. 1–36. http://dx.doi.org/10.1167/7.2.20.
36. Dobres, J., N. Chahine, B. Reimer, D. Gould, and N. Zhao. The Effects of Chinese Typeface Design, Stroke Weight, and Contrast Polarity on Glance-Based Legibility. *Displays,* Vol. 41, 2016, pp. 42–49. http://dx.doi.org/10.1016/j.displa.2015.12.001.
37. Sanford, E.C. The Relative Legibility of the Small Letters. *American Journal of Psychology,* Vol. 1, No. 3, 1888, pp. 402–435. http://dx.doi.org/10.2307/1411012.
38. Meyer, D.E., and R.W. Schvaneveldt. Facilitation in Recognizing Pairs of Words: Evidence of a Dependence Between Retrieval Operations. *Journal of Experimental Psychology,* Vol. 90, No. 2, 1971, pp. 227–234. http://dx.doi.org/10.1037/h0031564.
39. Medler, D.A., and J.R. Binder, Eds. *MCWord: An On-Line Orthographic Database of the English Language.* 2005. http://www.neuro.mcw.edu/mcword.
40. Levitt, H. Transformed Up–Down Methods in Psychoacoustics. *Journal of the Acoustical Society of America,* Vol. 49, No. 2B, 1971, pp. 467–477. http://dx.doi.org/10.1121/1.1912375.
41. Leek, M.R. Adaptive Procedures in Psychophysical Research. *Perception and Psychophysics,* Vol. 63, No. 8, 2001, pp. 1279–1292. http://dx.doi.org/10.3758/BF03194543.
42. R Core Team. *R: A Language and Environment for Statistical Computing.* The R Foundation for Statistical Computing, Vienna, Austria, 2016.
43. Bigelow, C., and S. Matteson. Font Improvements in Cockpit Displays and Their Relevance to Automobile Safety. Presented at the Society of Information Displays 2011 Vehicle Displays and Interfaces Symposium, University of Michigan, Dearborn, 2011.
44. Piepenbrock, C., S. Mayr, I. Mund, and A. Buchner. Positive Display Polarity Is Advantageous for Both Younger and Older Adults. *Ergonomics,* Vol. 56, No. 7, 2013, pp. 1116–1124. http://dx.doi.org/10.1080/00140139.2013.790485.
45. Piepenbrock, C., S. Mayr, and A. Buchner. Smaller Pupil Size and Better Proofreading Performance with Positive than with Negative Polarity Displays. *Ergonomics,* Vol. 57, No. 11, 2014, pp. 1670–1677. http://dx.doi.org/10.1080/00140139.2014.948496.
46. Buchner, A., and N. Baumgartner. Text–Background Polarity Affects Performance Irrespective of Ambient Illumination and Colour Contrast. *Ergonomics,* Vol. 50, No. 7, 2007, pp. 1036–1063. http://dx.doi.org/10.1080/00140130701306413.

The views and conclusions expressed are those of the authors and have not been sponsored, approved, or endorsed by Toyota or plaintiffs' class counsel.

The Standing Committee on Traffic Control Devices peer-reviewed this paper.