Bayesian adaptive assessment of the reading function for vision: The qReading method

Fang Hou

School of Ophthalmology & Optometry and Eye Hospital, Wenzhou Medical University, Wenzhou, Zhejiang, China ✉

Yukai Zhao

Center for Cognitive and Brain Sciences, Center for Cognitive and Behavioral Brain Imaging, and Department of Psychology, The Ohio State University, Columbus, OH, USA ✉

Luis Andres Lesmes

Adaptive Sensory Technology, Inc., San Diego, CA, USA ✉

Peter Bex

Department of Psychology, Northeastern University, Boston, MA, USA ✉

Deyue Yu

College of Optometry, The Ohio State University, Columbus, OH, USA ✉

Zhong-Lin Lu

Center for Cognitive and Brain Sciences, Center for Cognitive and Behavioral Brain Imaging, and Department of Psychology, The Ohio State University, Columbus, OH, USA ✉

Reading is a fundamental skill that can be significantly affected by visual disabilities. Reading performance, which typically is measured as reading speed with a reading chart, is a key endpoint for quantifying normal or abnormal vision. Despite its importance for clinical vision, existing reading tests for vision are time consuming and difficult to administer. Here, we propose a Bayesian adaptive method, the qReading method, for automated assessment of the reading speed versus print size function. We implemented the qReading method with a word/nonword lexical decision task and validated the method with computer simulations and a psychophysical experiment. Computer simulations showed that both the interrun standard deviation and intrarun half width of the 68.2% credible interval of the estimated reading speeds from the qReading method were less than 0.1 log10 units after 150 trials, with a bias of 0.05 log10 units. In the psychophysical experiment, reading functions measured by the qReading and Psi methods (Kontsevich & Tyler, 1999) in a word/nonword lexical decision task were compared. The estimated reading functions obtained with the qReading and Psi methods were highly correlated ($r = 0.966 \pm 0.004$, $p <$ 0.01). The precision of the qReading method with 225 trials was comparable to that of the Psi method with 450 trials. We conclude that the qReading method can precisely and accurately assess the reading function in much reduced time, with great promise in both basic research and clinical applications.

Introduction

In adulthood, reading is one of the most crucial daily activities for engaging and communicating with others and exchanging information. Impairment in reading can have a significant impact on the quality of life (Mitchell et al., 2008) and is often cited as a major factor in patients seeking professional help for eye-related problems (Elliott et al., 1997). Reading performance has also served as an outcome measure in clinical trials for assessing the effectiveness of treatments (Mahmood et al., 2015), surgical procedures (Jonker et al., 2015; Tang, Zhuang, & Liu, 2014), and rehabilitation techniques (Binns et al., 2012; Stelmack et al., 2017).

Because individual letters embedded in a word are more difficult to recognize than single letters (Whitney & Levi, 2011), reduced visual acuity cannot fully reveal reading impairment in a number of ophthalmic diseases (Crossland, Culham, & Rubin, 2005; Legge, Ross, Isenberg, & LaMay, 1992; McClure, Hart, Jackson, Stevenson, & Chakravarthy, 2000). On the other hand, reading speed, one of the most commonly used reading performance measures, is a strong predictor of visual ability and vision-related quality of life metrics for patients with vision loss (Hazel, Petre, Armstrong, Benson, & Frost, 2000; McClure et al., 2000). Reading speed has also been found to correlate with the pathological characteristics of age-related macular degeneration (Cacho, Dickinson, Smith, & Harper, 2010; Richter-Muecksch, Stur, Stifter, & Radner, 2006). All of these suggest that reading performance is an important component in evaluating functional vision.

There are several reading tests that measure reading acuity as well as reading speed at multiple print sizes, including Bailey–Lovie word reading charts (Bailey & Lovie, 1980), MNREAD charts (Precision Vision; see Ahn, Legge, & Luebker, 1995; Mansfield, Ahn, Ge, & Leubeker, 1993), SKread Charts (Precision Vision; see MacKeben, Nair, Walker, & Fletcher, 2015), and RADNER reading charts (Neumed AG, AT, Precision Vision; see Radner et al., 1998). The maximum reading speed and critical print size can be derived from the reading speed versus print size function. However, as Kingsnorth and Wolffsohn (2015) pointed out, chart-based reading tests can have learning effects and are cumbersome to use and too time consuming. They require an examiner to simultaneously present the reading material, take manual time measurements, and record reading accuracy. To obtain a reading speed versus print size function, examiners usually need to measure reading speeds at eight to ten print sizes, taking about 5–15 min in chart-based tests (Ahn et al., 1995; Bailey & Lovie, 1980; MacKeben et al., 2015; Mansfield et al., 1993; Radner et al., 1998), and an upper bound of 5 min per print size in a more precise test (Legge, Ross, Luebker, & LaMay, 1989). Moreover, deriving reading performance metrics from the raw data is also laborious (Kingsnorth & Wolffsohn, 2015) and susceptible to subjective factors (Cheung, Kallie, Legge, & Cheong, 2008).

There have been attempts to develop computerized reading tests (Dexl, Schlogel, Wolfbauer, & Grabner, 2010; Legge et al., 1989), including tests implemented on mobile devices (Calabrèse, Gamam, Mansfield, & Legge, 2014; Calabrèse et al., 2018; Kingsnorth & Wolffsohn, 2015). Although they have eliminated the requirement for the examiner to manually record reading time and automated data analysis, these tests still need an examiner to enter reading errors and can only measure reading speed at one print size at a time,

without considering any relationship between reading speeds at different print sizes.

There is a clear need for reading tests that are easy to administer, relatively less time-consuming, and can provide accurate and precise estimate of reading performance. In this study, we adopt the Bayesian adaptive testing strategy to develop a new reading test, the qReading method, that is fast, requires minimal examiner involvement, and has high accuracy and precision. The Bayesian adaptive testing strategy combines the Bayes rule and an information-theoretic framework to select the most informative stimulus in each trial and accumulate information about the reading speed versus print size function throughout the entire test procedure. The strategy has demonstrated great success in measuring a single sensory threshold (Kontsevich & Tyler, 1999) and can achieve even higher efficiency when it is applied to measure more complex visual functions by exploiting functional regularities in human behavior (Kujala & Lukka, 2006; Lesmes, Jeon, Lu, & Dosher, 2006; Lesmes, Lu, Baek, & Albright, 2010; Watson, 2017).

To develop qReading, we exploited a well-established functional relationship between reading speed and print size (Legge & Bigelow, 2011) to gain information of the entire reading speed versus print size function from each test trial at a single print size and presentation duration. In addition, we adopted the word/nonword lexical decision task to quantify a specific subtask of reading. Current reading tests require an examiner to judge reading accuracy and enter the number of reading mistakes. This renders them far less efficient than automated tasks that can be scored by a computer. However, automated computer scoring limits the specific reading tasks that can be implemented for reading testing. The word/nonword lexical decision task has been widely used in studies of word recognition (Meyer, Schvaneveldt, & Ruddy, 1975) and is considered as a probe to investigate the reading process (Cohen et al., 2000; Wandell, 2011) and an important assessment of reading ability (Gijsel, van Bon, & Bosman, 2004; Katz et al., 2012). In qReading with the lexical decision task, a letter string is briefly presented and followed by a mask, and the observer is asked to report if the letter string is a word or a nonword. The task can be considered as a special case (one-word version) of an rapid serial visual presentation (RSVP) reading task (Chung, Mansfield, & Legge, 1998). The reading speed in words per minute (wpm) can be computed as the reciprocal of threshold exposure duration (in seconds) times 60. As in many conventional reading tests for vision, we manipulate the exposure duration and print size of the stimuli in the lexical decision task to focus on visual factors in reading, keeping language comprehension factors minimal.

In the following, we describe the logic and core algorithm of the qReading method as well as computer simulations and a psychophysical experiment conducted to evaluate its performance. The results show that the qReading method can generate precise and accurate measures of reading speed versus print size function, and may provide a key endpoint for quantifying normal or abnormal vision.
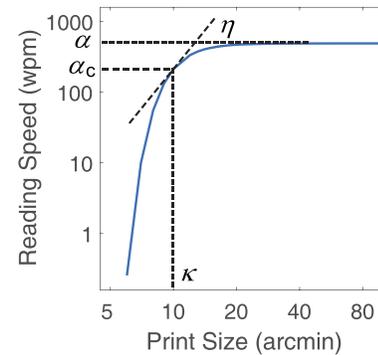
The qReading Method

It is well known that reading speed increases sharply with print size at small print and plateaus at large print sizes (Legge & Bigelow, 2011). The qReading method exploits this functional regularity between reading speed and print size to achieve greater efficiency. Specifically, the qReading method adopted an exponential-decay function form to describe the reading speed versus print size curve (Cheung et al., 2008):

$$\log 10(speed(size))$$
$$= \log10(\alpha) - (\log10(\alpha) - \log10(\alpha_C))$$
$$\times \exp\left(-\frac{(\log10(size) - \log10(\kappa))}{\eta}\right), \quad (1)$$

where $speed(size)$ is in wpm, α is the asymptote of reading speed in large print sizes, corresponding to the maximum reading speed, κ is the print size at which the reading speed is α_C wpm, and η controls the ascending rate of the reading speed versus print size function, α_C is defined as 360 wpm in the current study. Therefore, the reading speed versus print size function is characterized by three parameters: α, κ and η (Figure 1). Reading speed at a given print size is defined by the threshold exposure duration $\tau(size)$ (in seconds) with which the observer performs the lexical decision task at 80.3% correct in that print size:

$$speed(size) = 60/\tau(size). \quad (2)$$

The probability that the observer can perform the word/nonword lexical decision task correctly in a given print size and exposure duration condition is described by a psychometric function:

$$\Psi(duration(size)) = \gamma\lambda + (1 - \lambda)$$
$$\times \left(\gamma + (1 - \gamma)\left(1 - \exp\left(-\left(\frac{duration}{\tau(size)}\right)^{\beta}\right)\right)\right), \quad (3)$$

where $\tau(size)$ is the threshold exposure duration corresponding to 80.3% percent correct in the print size condition, $\gamma = 0.5$ is the guessing rate in the word/nonword lexical decision task, $\lambda = 0.04$ is the lapse rate for the task (Lesmes et al., 2010; Wichmann & Hill, 2001), and the slope β was set to 2.0 based on pilot data

Figure 1. The reading speed versus print size function is characterized by three parameters: α, the asymptotic reading speed in very large print sizes, κ, the print size at which the reading speed is 360 wpm, and η, the ascending rate of the exponential function. Both axes are in log10 units.

collected on human observers.[1] Combined together, Equations 1, 2, and 3 can model the response accuracy of the observer in any print size and exposure duration condition in the lexical decision task.

Before each test, the qReading method defines the parameter space for all the possible reading functions $\theta = (\alpha, \kappa, \eta)$ and a prior distribution of parameters $p_0(\theta)$ representing the experimenter's prior knowledge of the probability of different reading curves. It also defines the stimulus space that contains all possible print sizes and presentation durations to be tested in the experiment $x = (size, duration)$.

In each trial, the qReading method searches for the optimal print size and exposure duration via an information-theoretic approach, presents the optimal stimulus to the observer, and collects the observer's response. Figure 2 illustrates how the qReading works for a simulated observer. In Figure 2a, c, and e, the mutual information of all potential stimuli in the stimulus space is shown in Trials 3, 23, and 200. The location with the maximum mutual information, as indicated by the black square, represents the optimal stimulus used in the corresponding trial. The method accumulates information about the parameters α, κ, and η by updating their joint posterior distribution based on the observer's response via Bayes' rule. Because the performance of the observer in any print size and duration condition is jointly determined by these parameters, the qReading method can obtain information about the entire reading function in each trial instead of measuring reading speed at one print size at a time and therefore greatly improve test efficiency. As shown in Figure 2b, d, and f, the estimated reading speed versus print size curve is updated according to the responses made by observers (blue circles and red crosses). As trial number increases, the estimated reading speed versus print size function approaches the true function.
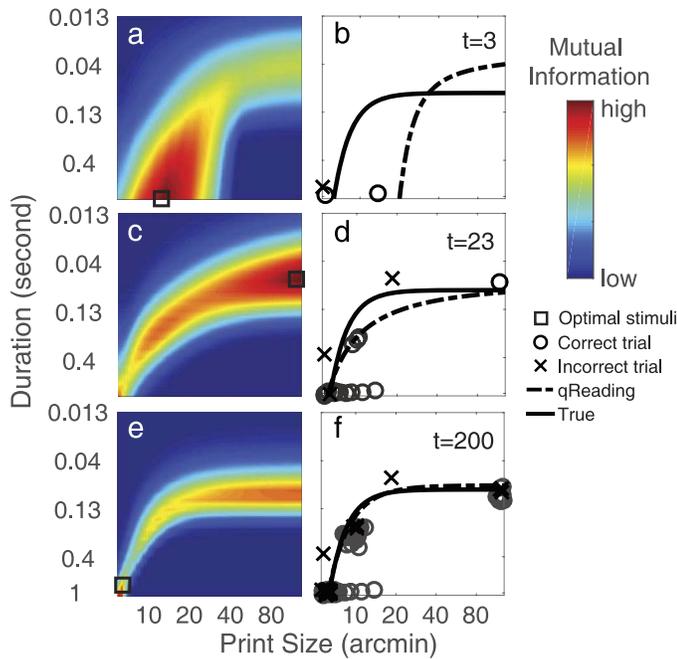
Figure 2. Illustrations of the qReading method in a simulation. (a, c, and e) Mutual information as a function of stimulus duration and print size in Trials 3, 23, and 200. The optimal stimulus, indicated by the square symbol, is selected to maximize the mutual information in Trials 3, 23, and 200, respectively. (b, d, and f) Stimuli placement in Trials 3, 23, and 200, respectively. Dashed curve: estimated reading speed versus print size function by the qReading method. Continuous curve: the true reading speed versus print size curve. Crosses: incorrect responses; Circles: correct responses. The locations of the crosses, circles, and squares indicate stimulus conditions.

The procedure is reiterated until a certain number of trials is reached. The estimated parameter values as well as the reading speed versus print size function can be derived from the posterior distributions of the parameters. For a more formal description of the qReading method, please see Appendix A.

Computer simulations

We first carried out computer simulations to test the qReading method. A simulated observer performed a two-alternative forced-choice word/nonword lexical decision task. The parameters of the simulated observer were: $\alpha_{true} = 1,556$ wpm, $\kappa_{true} = 9.26$ arcmin, and $\eta_{true} = 0.129$ log10 arcmin, based on the average parameter values obtained from a pilot experiment. The true reading speed versus print size function was then calculated using Equation 1, and used to generate the simulated observer's response probabilities using Equations 2 and 3. Then we used the qReading method

to estimate the reading function of the simulated observer based on her response in each trial.

We defined the following parameter space in the qReading method: 15 values evenly sampled from 2.55 to 3.75 (in log10 wpm units) for the asymptote α, 15 values evenly sampled from 0.46 to 1.66 (log10 arcmin units) for the critical size κ, and 15 values evenly sampled in log space from 0.079 to 1 (log10 arcmin units) for the ascending rate η. The prior of the parameters is defined as a uniform distribution in the corresponding region of the three-dimensional space. The range of possible stimuli is 60 print sizes from 5.79 to 129 arcmin and 20 durations from 0.013 to 1 s. The stimulus space was sampled evenly in log units in both dimensions.

The simulated observer was tested with 500 qReading runs. Each qReading run had 300 trials. In order to evaluate the performance of the qReading method, we examined the precision and bias of the estimated reading speed versus print size functions obtained from the qReading method.

The precision of a method can be gauged by the variability of its estimates. A smaller variability means a higher precision. In this study, the interrun standard deviation of the estimated parameters of the reading speed versus print size function was computed as:

$$SD_{inter} = \sqrt{\frac{\sum_{j=1}^{500}\left(\left(\log_{10}(v_j) - \log_{10}(\bar{v})\right)\right)^2}{500}}, \quad (4)$$

where v_j is the estimated parameters α, κ, or η, in the jth run, and $\bar{v}$ is the mean of v_j over 500 runs. The interrun standard deviation of the estimated reading speed was calculated as the mean standard deviation of the estimated reading speeds over all 60 print sizes.

We also examined the precision with intrarun variability, or the half width of the 68.2% credible interval (HWCI) of the posterior distribution of the parameters and the 68.2% HWCI of the distribution of the estimated reading speeds in a single qReading run (Clayton & Hills, 1993; Hou, Lesmes, Bex, Dorr, & Lu, 2015). The latter was performed with a resampling procedure. Five hundred sets of parameters θ were sampled from the posterior distribution $p_t(\theta)$ from a single qReading run. They were used to construct 500 reading functions and estimate the 68.2% HWCIs. The resampling procedure can take into account of the covariance structure in the posterior distribution of the reading speed versus print size function parameters (Hou et al., 2015).

The interrun standard deviation (blue) and intrarun HWCI (red) of the parameters α, κ, and η, as well as the estimated reading speeds from the qReading method, are plotted as functions of trial number in Figure 3a through d, respectively. Both interrun standard deviation and intrarun HWCI decreased rapidly in about 50
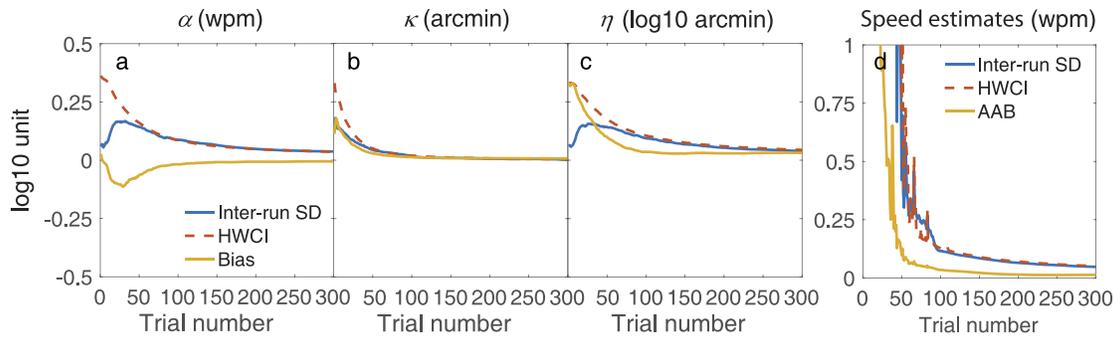
Figure 3. Interrun standard deviation, intrarun 68.2% HWCI, and bias of the estimated parameters α (in a), κ (in b), and η (in c) of the reading speed versus print size function of the simulated observer as functions of trial number. (d) Average interrun standard deviation, intrarun 68.2% HWCI, and bias of the estimated reading speed of the simulated observer as functions of trial number. Solid blue lines: interrun standard deviation; dashed red lines: intrarun HWCI; solid yellow lines: Bias (in a, b, and c) and AAB (in d).

trials. The interrun standard deviation for estimated α, κ, and η and reading speeds were 0.143, 0.044, 0.142, and 0.420 log10 units after 50 trials, respectively, and decreased to 0.065, 0.009, 0.068, and 0.084 log10 units after 150 trials, respectively. The intrarun HWCI for estimated α, κ, and η and reading speeds were 0.161, 0.051, 0.177, and 0.723 log10 units after 50 trials, respectively, and decreased to 0.060, 0.011, 0.075, and 0.090 log10 units after 150 trials, respectively.

The accuracy of a method quantifies how much the estimated value deviates from the truth. We computed the bias of the estimated parameters as

$$Bias = \frac{\sum_{j=1}^{500}\left(\log_{10}(v_j) - \log_{10}(v_{\text{true}})\right)}{500}, \quad (5)$$

where v_j is any one of the estimated parameters α, κ, or η in the jth run, and v_{true} is the true value of the parameter. We also computed the average absolute bias (AAB) of the estimated reading speeds:

$$AAB = \frac{\sum_{k=1}^{60}\left|\sum_{j=1}^{500}\left(\log_{10}\left(S_{j,k}\right) - \log_{10}\left(S_{\text{true},k}\right)\right)\right|}{500 \times 60}, \quad (6)$$

Where $S_{j,k}$ is the estimated speed at the kth print size in the jth run, and $S_{\text{true},k}$ is the true reading speed at the kth print size.

The bias of the estimated parameters α, κ, and η and the AAB of the estimated reading speeds from the qReading method are plotted as functions of trial number in Figure 3a through d, respectively. The bias of the estimated α, κ, and η and the AAB of the estimated speeds were −0.064, 0.030, 0.100, and 0.099 log10 units after 50 trials, respectively, and decreased to −0.008, 0.010, 0.029, and 0.022 log10 units after 150 trials, respectively. The simulation results showed clearly that the qReading method could deliver very precise and accurate assessment of the reading speed versus print size function efficiently.

Psychophysical experiment

To further validate the qReading method, we measured the reading speed versus print size function from four human observers with the qReading method and the Psi method (Kontsevich & Tyler, 1999) in a word/nonword lexical decision task. The Psi method was used to provide an independent measure of reading speed at each of a range of print sizes one at a time. The results obtained from the Psi method can be used to (a) test whether the reading speed versus print size functions estimated from the qReading method, which assesses the parameters of the entire function in each trial, is consistent with those obtained in a series of print sizes, (b) compare the two methods in terms of relative efficiency, and (c) test some of the underlying assumptions of the qReading method. The settings of the qReading method used in the experiment were almost the same as those used in the simulations except that the range and sampling of print sizes and presentation durations were adjusted to accommodate the physical limits (i.e., pixel size and refresh interval) of the monitor. The possible stimuli consisted of 50 print sizes from 4.34 to 89.7 arcmin evenly sampled in log units and 33 durations from 0.013 to 1 s evenly sampled in log units.

Methods

Observers

The second author (S1) and three other naive observers (S2–S4), aged 20 to 33 years, all with normal or corrected-to-normal vision, participated in the study. S2–S4 were native English speakers. The study was approved by the institutional review board of human subject research at The Ohio State University. Written consent was obtained before the experiment.
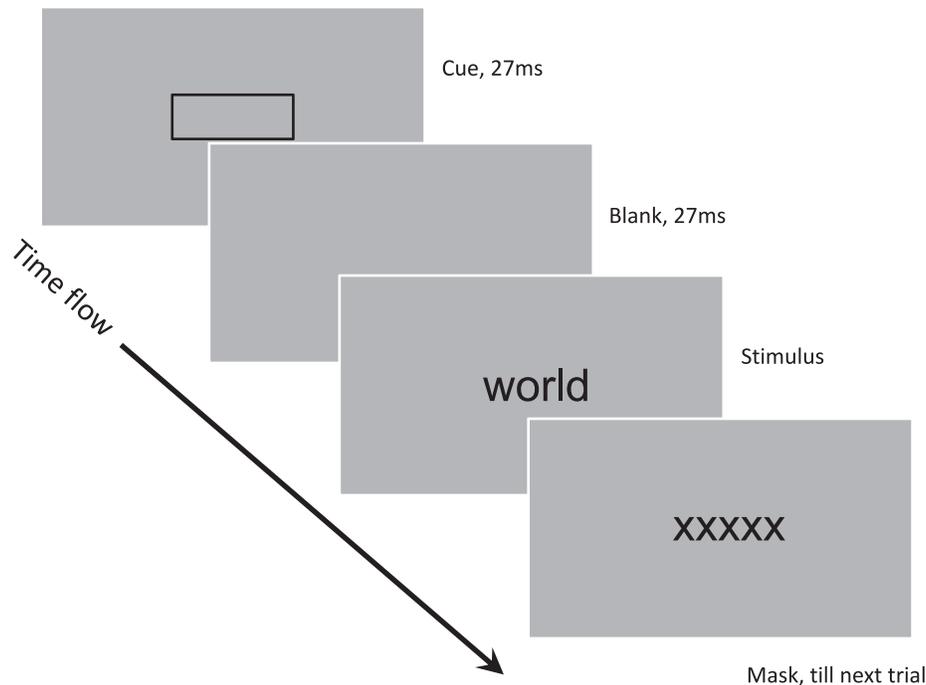
Figure 4. Illustration of the display sequence in one trial of the experiment.

Apparatus

All programs used in this study were written in MATLAB (MathWorks, Natick, MA) with Psychtoolbox extensions (Kleiner, Brainard, & Pelli, 2007) and run on a PC computer. Stimuli were displayed on a gamma-corrected Dell 19-in. M993s CRT monitor. The monitor had a 1,600 × 1,200 pixel resolution and a vertical refresh rate of 75 Hz. The mean luminance of the monitor was 33 cd/m². Observers viewed the stimuli binocularly at a distance of 106.6 cm in a dark room.

Stimuli

The stimuli used in the experiment were five-letter strings. The letter strings were presented in black on a gray background (33 cd/m²) in the Arial font style. Fifty print sizes (evenly sampled in log space from 4.34 to 89.7 arcmin) and 33 exposure durations (evenly sampled in log space from 0.013 to 1 s) were used in the study. The sizes and durations were rounded to the nearest physically available values in the unit of pixels and refresh intervals, separately. The Psi method was applied to measure threshold reading speeds at six print sizes that were selected for each observer based on data collected in the practice session.

Procedure

MCWord (Medler & Binder, 2005) was used to create five-letter word and nonword stimuli, which are generated from the CELEX English database and are based on the frequencies of written and spoken text from almost 18 million instances of word use. The most frequent 500 real five-letter words from the database were used as word stimuli, and 500 nonword stimuli were generated with constrained trigram statistics that match three letter combinations in the database. In each trial, a five-letter string was randomly selected from the pre-generated word/nonword pool. The string could be either a word or a nonword with equal probability. The observers had to decide whether the string is a word or nonword. Before the main experiment, each observer was given a practice session of 225 trials using the qReading procedure. Data from the practice session were used to determine the six print sizes to provide adequate sampling of the reading speed versus print size function in the subsequent Psi method test for each observer.

Each observer finished four experimental sessions, each lasting approximately 25 min and consisted of 900 trials (450 qReading and 450 Psi trials). Each session comprised two consecutive qReading runs (450 = 225 trials × 2 runs) randomly interleaved, trial-by-trial, with one Psi run that was used to estimate the threshold reading speed at six print sizes (450 = 75 trials × 6 sizes). The order of print sizes in the Psi procedure was random across trials.

The display sequence of one trial is illustrated in Figure 4. Each trial began with a 27-ms presentation of a rectangle box in the center of the display. The size of the box was the same as that of the outline of the to-be-presented letter string. This was followed by a 27-ms blank screen, a five-letter string with a certain print size
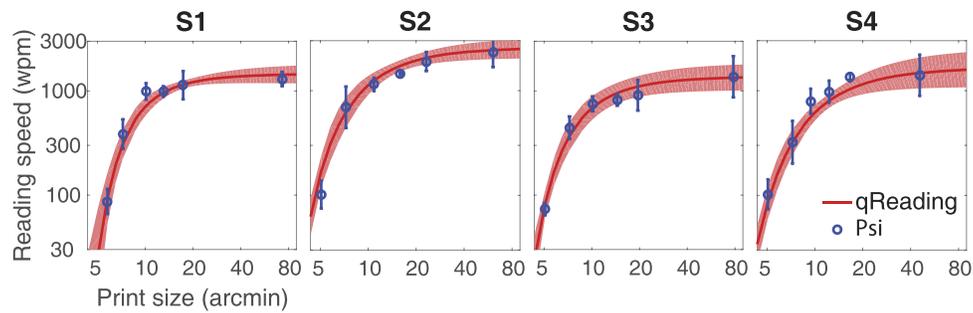
Figure 5. Estimated reading functions from the qReading and Psi methods for each observer. The curves represent the mean reading curves from eight qReading runs. Circles represent the mean estimates from four Psi runs. The shaded area and error bars represent ±1 *SD*.

and exposure duration, and a mask made of "xxxxx" in the same print size that was present until the next trial started. For the qReading trials, the print size and exposure duration of the stimulus were determined by the qReading algorithm. For the Psi trials, the print size was randomly selected from the predetermined sizes and the exposure duration was chosen by the Psi method. Observers were instructed to use the two buttons on the mouse to indicate if the letter string was a word or nonword. No feedback was provided. A new trial started 500 ms after the response.

Results

Agreement between the qReading and Psi methods

Figure 5 presents the reading speed versus print size functions of four observers, obtained with the qReading and Psi methods. The estimated reading speed versus print size functions obtained from the qReading method are shown as solid curves, and those obtained from the Psi method are shown as circles. The standard deviation of the estimated reading speeds from the qReading measurements was calculated from eight repeated runs. The standard deviation of the Psi estimates was computed from four repeated runs.

Inspection of the estimates from both methods suggests excellent agreement.

We compared the parameters estimated from the two methods. Equation 1 was fitted to the six reading speeds versus print size curve estimated by the Psi method in each session for each observer. The estimated reading parameters, α, κ, and η from the Psi method, were determined by the best-fitted parameter values, averaged across sessions and plotted against the average parameters measured directly from the qReading method in Figure 6a through c. The Pearson correlation coefficients between the estimated parameters from the two methods were 0.948 ($p = 0.052$), 0.997 ($p = 0.003$) and 0.975 ($p = 0.025$) for α, κ, and η, respectively. No significant difference was found between the estimated α, $t(3) = 2.76$, $p = 0.067$, and κ, $t(3) = 1.97$, $p = 0.144$, from the two methods. The estimated η from the qReading method was significantly smaller than that from the Psi method, $t(3) = 4.37$, $p = 0.022$.

In Figure 6d, estimated reading speeds at the six print sizes used in the Psi procedure from both the qReading and Psi methods are plotted against each other. Again, data from the two methods showed excellent agreement. Because an exponential reading speed versus print size function with three parameters
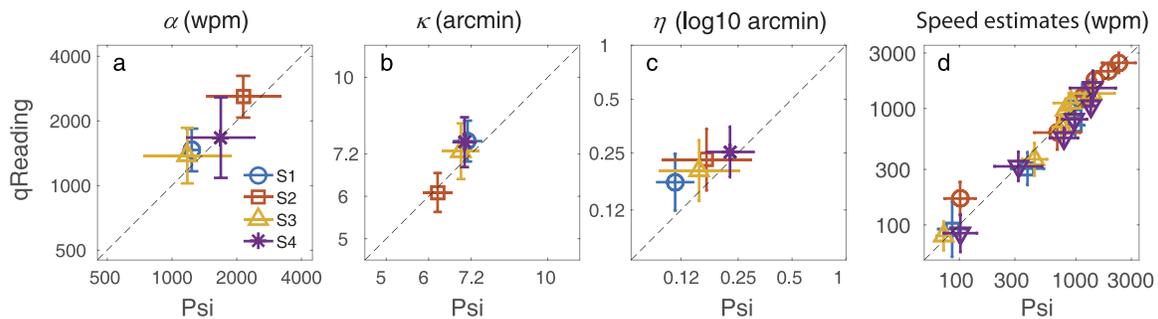

Figure 6. The estimated parameters of the reading speed versus print size function α, κ and η as well as the estimated reading speeds from the qReading and Psi method are plotted against each other in Panels (a) through (d), respectively. Four different symbols represent data from the four observers.
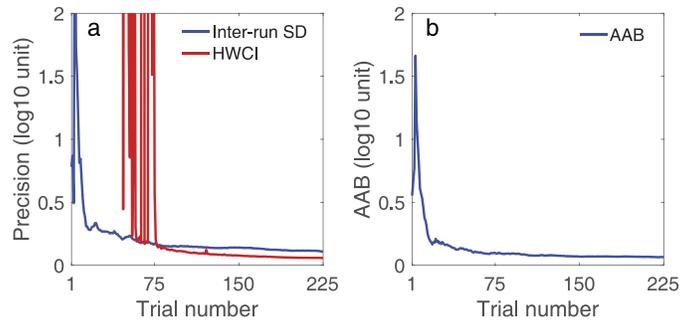
Figure 7. (a) The interrun standard deviation and intrarun HWCI of the estimated reading speed from the qReading method as functions of trial number. Blue line: interrun standard deviation; red line: intrarun HWCI. (b) The AAB of the estimated reading speed from the qReading method as a function of trial number.

was used (Equation 1), the reading speeds obtained from the qReading method are not independent across print size conditions. To compute the correlation of the reading speeds estimated by the two methods, we carried out a special procedure for each observer to eliminate the dependency across print size conditions (see Appendix B for the detailed procedure). Across all observers, the average correlation coefficient between the estimated reading speeds obtained with the two methods was 0.969 ± 0.005 ($p < 0.01$ for all observers). A repeated-measure ANOVA with print size and method as factors was conducted on the estimated reading speeds obtained from the two methods. Method had no significant effect, $F(1, 15) = 0.036$, $p = 0.862$.

Precision of the qReading method

We computed the average interrun standard deviation and intrarun 68.2% HWCI of the estimated reading speed across print sizes and observers from the eight qReading runs (Figure 7a). The average interrun standard deviation of the estimated reading speed from the qReading procedure was 0.172 ± 0.077, 0.141 ± 0.048, and 0.109 ± 0.044 log 10 units after 75, 150, and 225 trials, respectively. The average 68.2% HWCI of the estimated reading speed from the qReading procedure was 0.570 ± 4.00, 0.078 ± 0.077, and 0.058 ± 0.040 log10 units after 75, 150, and 225 trials, respectively. For comparison, the average interrun standard deviation and HWCI of the estimated reading speed measured by the Psi method was 0.112 ± 0.055 and 0.067 ± 0.08 log10 units with 450 trials. The standard deviation and 68.2% HWCI of the estimated reading speeds from the qReading method decreased rapidly in the beginning of the procedure. After about 80 trials, the HWCI was almost the same as the interrun standard deviation.

Using the mean reading speeds from the four Psi runs as the "truth," we computed the average absolute bias of the estimated reading speed obtained with the qReading method using the following equation:

$$AAB = \frac{\sum_{k=1}^{50} \left| \sum_{j=1}^{8} \left(\log_{10}\left(S_{j,k}\right) - \log_{10}\left(S_{\text{true},k}\right) \right) \right|}{50 \times 8}. \quad (7)$$

The average absolute bias of the estimated reading speed obtained with the qReading procedure across print sizes and observers is plotted as a function of trial number in Figure 7b. The AAB was 0.098 ± 0.026, 0.069 ± 0.027, and 0.065 ± 0.012, log10 units after 75, 150, and 225 trials, respectively. Given that the "true" reading speeds estimated by the Psi method had an average standard deviation of 0.11 log10 units, an average absolute bias of 0.065 log10 units is relatively small.

Test–retest reliability

A reliable measurement should not only produce the same ranking upon the same respondents but also produce exactly the same values in different measurements. In order to examine the test–retest reliability of the qReading method, we analyzed the overall concordance correlation coefficient (OCCC; see Barnhart, Haber, & Song, 2002; Lin, 1989, 2000) for assessing agreement among eight qReading measurements. The OCCC is the weighted average of the pair-wise concordance correlation coefficient between any two qReading measurements, which measures the agreement between two tests by measuring the variation from the 45° line (diagonal) through the origin. The mean OCCC of the estimated reading speed across four observers was 0.891 ± 0.024.

Assumption check

One underlying assumption of the qReading method is that the slope β of the psychometric function for the word/nonword lexical decision task is the same at

Observer	S1	S2	S3	S4
Model I				
$\chi^2(48)$	51.5	37.5	37.3	48.0
p	0.338	0.863	0.867	0.471
Model II				
$\chi^2(53)$	53.0	45.0	40.8	51.8
p	0.472	0.775	0.889	0.522
Model II vs. Model I				
$\chi^2(5)$	1.52	7.50	3.50	3.74
p	0.911	0.186	0.623	0.588

Table 1. Results of the model fit for all observers.

different print sizes and can be fixed at 2.0 (Equation 3). It is important to know if the fixed slope assumption is valid in the real human experiment and if the true slope differs from the assumed value of 2.0.

For each observer, data from the Psi method in all four sessions were pooled together. There were 300 trials in each print size condition, which were binned by dividing the log exposure duration into 10 equally spaced intervals. The percent of correct responses in the 10 bins allowed us to construct a raw psychometric function in each print size condition (Hou et al., 2010). Then we constructed two models (I and II) and fitted both models to the raw psychometric function using a maximum likelihood procedure. Model I, in which the psychometric functions of the word/nonword lexical decision task have different thresholds and slopes in different print size conditions, fit the raw psychometric functions well for all observers (χ^2 test, all ps > 0.05, Table 1). Model II, in which the thresholds are different but the slope is the same in different print size conditions, also provided good fits to the data for all observers (χ^2 test, all ps > 0.05, Table 1). A nested model test showed that Model II is statistically equivalent to Model I (χ^2 test, all ps < 0.05, Table 1), indicating that the fix slope assumption in the qReading method held true in our experimental data. From the best fitting Model II of each observer, we computed the averaged slope across observers. It was 2.06 ± 0.381, not significantly different from the assumed value in the qReading method, $t(3) = 0.312$, $p = 0.776$.

Discussion

The goal of the current study is to develop an automated and efficient psychophysical procedure to measure reading speed. Combining the Bayesian adaptive method and a word/nonword lexical decision task, we developed the qReading method that retains the precision of psychophysical testing but enables shorter testing time. Computer simulations showed that both the interrun standard deviation and intrarun

68.2% HWCI of the estimated reading speed from the qReading method were less than 0.1 log10 units with only 150 trials, with a bias of 0.05 log10 units. The results of a psychophysics experiment were largely consistent with the simulations. The estimated parameter of the reading function α, κ, and η, as well as reading speeds from the qReading method, were highly correlated with those from the Psi method. The interrun standard deviation, intrarun HWCI, and average absolute bias of the estimated reading speeds from the qReading method were 0.109 ± 0.045, 0.058 ± 0.040, and 0.065 ± 0.012 log 10 units after 225 trials, respectively. Moreover, to achieve the same amount of precision, the qReading method only needs about half the number of trials of the Psi method. The test–retest reliability, as indicated by the OCCC of the reading speed measured by the qReading method was 0.891 ± 0.024.

Although the parameters of the reading function correlated well between two methods, the decay constant η estimated from the qReading method was smaller than that derived from the Psi method, $t(3) = 4.37$, $p = 0.022$. However, it should not be considered as an inaccuracy issue for qReading method. In our experiment, reading speed was only sampled at six print sizes in the Psi test. It is possible that the number of data points was not sufficient to obtain an accurate estimate of η from the Psi test.

We also checked the assumptions underlying the qReading method. In the qReading method, we modeled the word/nonword lexical decision task with a two-alternative forced-choice Weibull psychometric function (Equation 3) with a fixed slope of 2.0 across different print sizes. Violation of these assumptions could substantially undermine the validity of the measurements from the qReading method. As shown by our analysis of the experimental data, the assumptions held well in the current study.

The qReading method inherits several advantages of the Bayesian adaptive testing framework. First, high efficiency was achieved through optimal stimulus selection. The qReading method searches the entire stimulus space for a stimulus condition that could provide the maximum expected mutual information (Kontsevich & Tyler, 1999; Kujala & Lukka, 2006). The efficiency of the qReading test was evidently supported by the rapidly decreasing standard deviation and bias in both the simulation and psychophysics studies (Figures 3 and 7). We used a one-step-ahead search procedure in the current implementation of the qReading method. Although this one-step-ahead greedy search algorithm provides asymptotically optimal results in most of the cases (Sims & Pelli, 1987), multiple-step-ahead search or global optimization can be used to gain extra efficiency when applicable (Kim, Pitt, Lu, & Myung, 2016).

The second advantage came from the Bayesian framework. The information accumulated through the posterior distribution during the test allowed us to gauge the variability of the estimated properties of the reading speed versus print size function in a single run of the qReading test. The intrarun variability represented by the 68.2% HWCI overlapped with interrun standard deviation after completing enough trials in the procedure (Figures 3 and 7). Similar findings were reported in a study on qCSF (Hou et al., 2015). The posterior distribution can also be exploited to test and detect reading performance changes affected by visual disease and/or other factors for an individual in different conditions. We have demonstrated the use of the posterior distribution in detecting CSF changes with the qCSF method (Hou et al., 2016). The usefulness of the qReading method in detecting changes of reading performance will be evaluated in future studies.

The qReading method also makes use of the well-established functional relationship between reading speed and print size (Legge & Bigelow, 2011). Instead of measuring reading speed at one print size at a time, the qReading method uses the information from observer's performance in a single trial, that is, in a single print size and exposure duration condition, to infer the properties of the entire reading speed versus print size function. The use of functional regularities therefore greatly improves the testing efficiency. It has been reported reading speed depends on other factors—for example, stimulus contrast (Legge, Rubin, & Luebker, 1987), eccentricity (Chung et al., 1998), and spacing between letters or words (Chung, 2004; Yu, Cheung, Legge, & Chung, 2007). These regularities could also be used by the qReading method to assess reading speed under multiple test conditions.

Another improvement came from the automated task. We have used a relatively simple word/nonword lexical decision task that does not require the experimenter to score the responses of the observers to probe the initial stage of reading process (Cohen et al., 2000; Wandell, 2011). As performance in the lexical decision task, mostly measured with reaction time, can be affected by linguistic factors such as word frequency and type of nonwords (Ratcliff, Gomez, & McKoon, 2004), we used the most frequent five-letter words and presented word or nonword stimuli with equal possibility to keep the language comprehension demand minimal. Instead of measuring reaction time, we manipulated the exposure duration and print size of the stimuli in the lexical decision task and measured response accuracy to focus on visual factors in reading.

The word/nonword lexical decision task used in the current implementation of the qReading method is related to paradigms used in previous reading tests in many ways. It can be considered as a one-word version of the RSVP silent reading test (Chung et al., 1998; Rubin & Turano, 1994) with unrelated words. In fact, unrelated random words have been extensively used in existing reading tests, such as the Bailey–Lovie word reading chart (Bailey & Lovie, 1980) and the SKread charts (Precision Vision; MacKeben et al., 2015). In this case, linguistic aspects such as grammar and context involving complex nonvisual factors are minimized, and observer's reading performance depends on word recognition alone (Radner, 2017).

The estimated reading speed in the current study is much greater than that reported in the literature (Calabrèse et al., 2016). This discrepancy may be due to different test settings as well as the frequency and length of the words used in different studies (Radner et al., 2002). The reading speed measured using the RSVP paradigm could be 2–4 times faster than that measured with static presentations (Rubin & Turano, 1994). Silent reading is generally faster than reading aloud (Lovie-Kitchin, Bowers, & Woods, 2000). In this study, high frequency words with a length of five were used. The high frequency and short length of the words may have also contributed to the fast reading speed (Radner et al., 2002).

The qReading method implemented with the word/nonword classification task in the current study is our first attempt to use the method to quantify a specific subtask of reading. The qReading method can be used with other reading materials and/or assessment tasks, provided that the assumptions on the psychometric properties of reading performance are valid. In a parallel study, we have also tested the method in an RSVP sentence-reading task in peripheral vision and found similar promising results (Shepard et al., 2017). In another study (Arango et al., 2017), we compared a computer-based sentence reading MNREAD test (Mansfield et al., 1993) with two RSVP reading tasks that used the qReading algorithm to control print size and presentation duration: five-letter word versus nonword judgement, and true or false judgement of four-word sentences (Crossland, Legge, & Dakin, 2008).

This paper is the first step in the development of the qReading method. We used high frequency words with a fixed length of five letters. Bailey and Lovie (1980) found that word length can affect the readability of words in patients with age-related macular degeneration. Words with different lengths can be used in future implementations of the qReading method to capture the characteristics of visual impairments. In ongoing and future work, we will continue to develop the qReading method and expand its application to a broad range of reading test paradigms. The new developments and applications will depend on the principles and statistical validation established in this paper.

Because the most common functional endpoint, visual acuity, is insensitive to many retinopathies, especially in the early stages of diseases where visual acuity may be within normal limits (Crossland et al., 2005), the reading function may provide important quantification of visual impairment for patient classification, and may be used for screening and monitoring the progression or remediation of visual impairment. The development of the qReading method is just a beginning of many future studies. As a powerful and versatile approach to measure reading performance, we expect many new developments and applications of the method in the future.

Keywords: reading, Bayesian adaptive test, precision, efficiency

Acknowledgments

Commercial relationships: none.
Corresponding author: Zhong-Lin Lu.
Email: lu.535@osu.edu.
Address: Center for Cognitive and Brain Sciences, Center for Cognitive and Behavioral Brain Imaging, and Department of Psychology, The Ohio State University, Columbus, OH, USA.

Footnotes

[1] In the current implementation of the qReading method, we assume that the shape of the underlying psychometric function is known. It has been demonstrated that the shape of the psychometric function was largely invariant in contrast detection (Chen et al., 2014; Hou et al., 2010) and letter identification (Hou, Lu, & Huang, 2014). This assumption was tested and validated in the psychophysical experiment in this study.

[2] [In Appendix A] The Psi method (Kontsevich & Tyler, 1999) is based on minimizing expected entropy, which is equivalent to maximizing the expected change of entropy $H(\theta) - H(\theta|r)$ (Kujala & Lukka, 2006; MacKay, 1992). Expected change of entropy is also known as mutual information $I(\theta|r) = H(\theta) - H(\theta|r)$ (Cover & Thomas, 1991). $I(\theta; r)$ represents the amount of information about θ given the observation of r. The mutual information is symmetric (Cover & Thomas, 1991): $I_{t-1}(\theta; r) = I_{t-1}(r; \theta)$. Also see (Kujala & Lukka, 2006).

References

Ahn, S. J., Legge, G. E., & Luebker, A. (1995). Printed cards for measuring low-vision reading speed. *Vision Research, 35*(13), 1939–1944.

Arango, T., Hou, F., Lesmes, L., Yu, D., Lu, Z. L., & Bex, P. (2017). Different reading tasks measure different reading behaviors. *Journal of Vision, 17*(10):1033, https://doi.org/10.1167/17.10.1033. [Abstract]

Bailey, I. L., & Lovie, J. E. (1980). The design and use of a new near-vision chart. *American Journal of Optometry and Physiological Optics, 57*(6), 378–387.

Barnhart, H. X., Haber, M., & Song, J. (2002). Overall concordance correlation coefficient for evaluating agreement among multiple observers. *Biometrics, 58*(4), 1020–1027.

Binns, A. M., Bunce, C., Dickinson, C., Harper, R., Tudor-Edwards, R., Woodhouse, M., . . . Margrain, T. H. (2012). How effective is low vision service provision? A systematic review. *Survey of Ophthalmology, 57*(1), 34–65, https://doi.org/10.1016/j.survophthal.2011.06.006.

Cacho, I., Dickinson, C. M., Smith, H. J., & Harper, R. A. (2010). Clinical impairment measures and reading performance in a large age-related macular degeneration group. *Optometry and Vision Science, 87*(5), 344–349, https://doi.org/10.1097/OPX.0b013e3181d9515c.

Calabrèse, A., Cheong, A. M., Cheung, S. H., He, Y., Kwon, M., Mansfield, J. S., . . . Legge, G. E. (2016). Baseline MNREAD measures for normally sighted subjects from childhood to old age. *Investigative Ophthalmology & Vision Science, 57*(8), 3836–3843, https://doi.org/10.1167/iovs.16-19580.

Calabrèse, A., Gamam, S., Mansfield, J. S., & Legge, G. E. (2014). Implementing the MNREAD reading acuity test on an iPad3. *Investigative Ophthalmology & Visual Science, 55*(13), 5601.

Calabrèse, A., To, L., He, Y., Berkholtz, E., Rafian, P., & Legge, G. E. (2018). Comparing performance on the MNREAD iPad application with the

MNREAD acuity chart. *Journal of Vision*, *18*(1):8, 1–11, https://doi.org/10.1167/18.1.8. [PubMed] [Article]

Chen, G., Hou, F., Yan, F. F., Zhang, P., Xi, J., Zhou, Y., . . . Huang, C. B. (2014). Noise provides new insights on contrast sensitivity function. *PLoS One*, *9*(3), e90579, https://doi.org/10.1371/journal.pone.0090579.

Cheung, S.-H., Kallie, C. S., Legge, G. E., & Cheong, A. M. Y. (2008). Nonlinear mixed-effects modeling of MNREAD data. *Investigative Ophthalmology & Visual Science*, *49*(2), 828–835, https://doi.org/10.1167/iovs.07-0555.

Chung, S. T. (2004). Reading speed benefits from increased vertical word spacing in normal peripheral vision. *Optometry & Vision Science*, *81*(7), 525–535.

Chung, S. T., Mansfield, J. S., & Legge, G. E. (1998). Psychophysics of reading. XVIII. The effect of print size on reading speed in normal peripheral vision. *Vision Research*, *38*(19), 2949–2962.

Clayton, D., & Hills, M. (1993). *Statistical models in epidemiology*. Oxford, UK: Oxford University Press.

Cohen, L., Dehaene, S., Naccache, L., Lehericy, S., Dehaene-Lambertz, G., Henaff, M. A., & Michel, F. (2000). The visual word form area: Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain*, *123*(Pt 2), 291–307.

Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory* (1st ed.). New York: Wiley.

Crossland, M. D., Culham, L. E., & Rubin, G. S. (2005). Predicting reading fluency in patients with macular disease. *Optometry & Vision Science*, *82*(1), 11–17.

Crossland, M. D., Legge, G. E., & Dakin, S. C. (2008). The development of an automated sentence generator for the assessment of reading speed. *Behavioral and Brain Functions*, *4*: 14, https://doi.org/10.1186/1744-9081-4-14.

Dexl, A. K., Schlogel, H., Wolfbauer, M., & Grabner, G. (2010). Device for improving quantification of reading acuity and reading speed. *Journal of Refractive Surgery*, *26*(9), 682–688, https://doi.org/10.3928/1081597x-20091119-01.

Elliott, D. B., Trukolo-Ilic, M., Strong, J. G., Pace, R., Plotkin, A., & Bevers, P. (1997). Demographic characteristics of the vision-disabled elderly. *Investigative Ophthalmology & Vision Science*, *38*(12), 2566–2575.

Gijsel, M. A. R., van Bon, W. H. J., & Bosman, A. M. T. (2004). Assessing reading skills by means of paper-and-pencil lexical decision: Issues of reliability, repetition, and word-pseudoword ratio. *Reading and Writing*, *17*(5), 517–536, https://doi.org/10.1023/B:READ.0000044599.98083.d8.

Hazel, C. A., Petre, K. L., Armstrong, R. A., Benson, M. T., & Frost, N. A. (2000). Visual function and subjective quality of life compared in subjects with acquired macular disease. *Investigative Ophthalmology & Vision Science*, *41*(6), 1309–1315.

Hou, F., Huang, C. B., Lesmes, L., Feng, L. X., Tao, L., Zhou, Y. F., & Lu, Z. L. (2010). qCSF in clinical application: Efficient characterization and classification of contrast sensitivity functions in amblyopia. *Investigative Ophthalmology & Vision Science*, *51*(10), 53655377, https://doi.org/10.1167/iovs.10-5468.

Hou, F., Lesmes, L., Bex, P., Dorr, M., & Lu, Z. L. (2015). Using 10AFC to further improve the efficiency of the quick CSF method. *Journal of Vision*, *15*(9):2, 1–18, https://doi.org/10.1167/15.9.2. [PubMed] [Article]

Hou, F., Lesmes, L. A., Kim, W., Gu, H., Pitt, M. A., Myung, J. I., & Lu, Z. L. (2016). Evaluating the performance of the quick CSF method in detecting contrast sensitivity function changes. *Journal of Vision*, *16*(6):18, 1–19, https://doi.org/10.1167/16.6.18. [PubMed] [Article]

Hou, F., Lu, Z. L., & Huang, C. B. (2014). The external noise normalized gain profile of spatial vision. *Journal of Vision*, *14*(13):9, 1–14, https://doi.org/10.1167/14.13.9. [PubMed] [Article]

Jonker, S. M., Bauer, N. J., Makhotkina, N. Y., Berendschot, T. T., van den Biggelaar, F. J., & Nuijts, R. M. (2015). Comparison of a trifocal intraocular lens with a +3.0 D bifocal IOL: Results of a prospective randomized clinical trial. *Journal of Cataract & Refractive Surgery*, *41*(8), 1631–1640, https://doi.org/10.1016/j.jcrs.2015.08.011.

Katz, L., Brancazio, L., Irwin, J., Katz, S., Magnuson, J., & Whalen, D. H. (2012). What lexical decision and naming tell us about reading. *Reading and Writing*, *25*(6), 1259–1282, https://doi.org/10.1007/s11145-011-9316-9.

Kim, W., Pitt, M. A., Lu, Z. L., & Myung, J. I. (2016). Planning beyond the next trial in adaptive experiments: A dynamic programming approach. *Cognitive Science*, *41*(8), 2234–2252, https://doi.org/10.1111/cogs.12467.

Kingsnorth, A., & Wolffsohn, J. S. (2015). Mobile app reading speed test. *British Journal of Ophthalmology*, *99*(4), 536–539, https://doi.org/10.1136/bjophthalmol-2014-305818.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception*, *36*, 14.

Kontsevich, L. L., & Tyler, C. W. (1999). Bayesian adaptive estimation of psychometric slope and threshold. *Vision Research*, *39*(16), 2729–2737.

Kujala, J. V., & Lukka, T. J. (2006). Bayesian adaptive estimation: The next dimension. *Journal of Mathematical Psychology*, *50*(4), 369–389, http://doi.org/10.1016/j.jmp.2005.12.005.

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, 1–22, https://doi.org/10.1167/11.5.8. [PubMed] [Article]

Legge, G. E., Ross, J. A., Isenberg, L. M., & LaMay, J. M. (1992). Psychophysics of reading. Clinical predictors of low-vision reading speed. *Investigative Ophthalmology & Vision Science*, *33*(3), 677–687.

Legge, G. E., Ross, J. A., Luebker, A., & LaMay, J. M. (1989). Psychophysics of reading. VIII. The Minnesota Low-Vision Reading Test. *Optometry & Vision Science*, *66*(12), 843–853.

Legge, G. E., Rubin, G. S., & Luebker, A. (1987). Psychophysics of reading—V. The role of contrast in normal vision. *Vision Research*, *27*(7), 1165–1177.

Lesmes, L. A., Jeon, S. T., Lu, Z. L., & Dosher, B. A. (2006). Bayesian adaptive estimation of threshold versus contrast external noise functions: The quick TvC method. *Vision Research*, *46*(19), 3160–3176, https://doi.org/10.1016/j.visres.2006.04.022.

Lesmes, L. A., Lu, Z. L., Baek, J., & Albright, T. D. (2010). Bayesian adaptive estimation of the contrast sensitivity function: The quick CSF method. *Journal of Vision*, *10*(3):17, 11–21, https://doi.org/10.1167/10.3.17. [PubMed] [Article]

Lin, L. I. (1989). A concordance correlation-coefficient to evaluate reproducibility. *Biometrics*, *45*(1), 255–268, https://doi.org/10.2307/2532051.

Lin, L. I. (2000). A note on the concordance correlation coefficient. *Biometrics*, *56*(1), 324–325.

Lovie-Kitchin, J. E., Bowers, A. R., & Woods, R. L. (2000). Oral and silent reading performance with macular degeneration. *Ophthalmic and Physiological Optics*, *20*(5), 360–370, https://doi.org/10.1016/S0275-5408(99)00088-5.

MacKay, D. J. C. (1992). Information-based objective functions for active data selection. *Neural Computation*, *4*(4), 590–604, https://doi.org/10.1162/neco.1992.4.4.590.

MacKeben, M., Nair, U. K., Walker, L. L., & Fletcher, D. C. (2015). Random word recognition chart helps scotoma assessment in low vision. *Optometry & Vision Science*, *92*(4), 421–428, https://doi.org/10.1097/OPX.0000000000000548.

Mahmood, S., Roberts, S. A., Aslam, T. M., Parkes, J., Barugh, K., Bishop, P. N., & Group, G. S. (2015). Routine versus as-needed Bevacizumab with 12-weekly assessment intervals for neovascular age-related macular degeneration: 92-week results of the GMAN trial. *Ophthalmology*, *122*(7), 1348–1355, https://doi.org/10.1016/j.ophtha.2015.03.017.

Mansfield, J., Ahn, S., Ge, L., & Leubeker, A. (1993). A new reading acuity chart for normal and low vision. *Noninvasive Assessment of the Visual System Technical Digest* (Vol. 3, pp. 232–235). Washington, DC.

McClure, M. E., Hart, P. M., Jackson, A. J., Stevenson, M. R., & Chakravarthy, U. (2000). Macular degeneration: Do conventional measurements of impaired visual function equate with visual disability? *British Journal of Ophthalmology*, *84*(3), 244–250.

Medler, D. A., & Binder, J. R. (2005). MCWord: An on-line orthographic database of the English language. Retrieved from http://www.neuro.mcw.edu/mcword/

Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. (1975). Loci of contextual effects on visual word recognition. In P. Rabbitt & S. Dornic (Eds.), *Attention and performance V* (pp. 98–118). London: Academic Press.

Mitchell, J., Wolffsohn, J., Woodcock, A., Anderson, S. J., Ffytche, T., Rubinstein, M., . . . Bradley, C. (2008). The MacDQoL individualized measure of the impact of macular degeneration on quality of life: Reliability and responsiveness. *American Journal of Ophthalmology*, *146*(3), 447–454, https://doi.org/10.1016/j.ajo.2008.04.031.

Radner, W. (2017). Reading charts in ophthalmology. *Graefe's Archive for Clinical and Experimental Ophthalmology*, *255*(8), 1465–1482, https://doi.org/10.1007/s00417-017-3659-0.

Radner, W., Obermayer, W., Richter-Mueksch, S., Willinger, U., Velikay-Parel, M., & Eisenwort, B. (2002). The validity and reliability of short German sentences for measuring reading speed. *Graefe's Archive for Clinical and Experimental Ophthalmology*, *240*(6), 461–467, https://doi.org/10.1007/s00417-002-0443-5.

Radner, W., Willinger, U., Obermayer, W., Mudrich, C., Velikay-Parel, M., & Eisenwort, B. (1998). [A new reading chart for simultaneous determination of reading vision and reading speed]. *Klinische*

Monatsblätter für Augenheilkunde, *213*(3), 174–181, https://doi.org/10.1055/s-2008-1034969.

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review*, *111*(1), 159–182, https://doi.org/10.1037/0033-295X.111.1.159.

Richter-Muecksch, S., Stur, M., Stifter, E., & Radner, W. (2006). Differences in reading performance of patients with Drusen maculopathy and subretinal fibrosis after CNV. *Graefe's Archive for Clinical and Experimental Ophthalmology*, *244*(2), 154–162, https://doi.org/10.1007/s00417-005-0063-y.

Rubin, G. S., & Turano, K. (1994). Low vision reading with sequential word presentation. *Vision Research*, *34*(13), 1723–1733, https://doi.org/10.1016/0042-6989(94)90129-5.

Shepard, T. G., Hou, F., Bex, P. J., Lesmes, L. A., Lu, Z. L., & Yu, D. (2017). The quick reading method: Its efficiency and accuracy in assessing reading performance in the periphery. *Investigative Ophthalmology & Visual Science*, *58*(8), 3278.

Sims, J. A., & Pelli, D. G. (1987). The ideal psychometric procedure. *Investigative Ophthalmology and Visual Science (Supplement)*, *28*, 366.

Stelmack, J. A., Tang, X., Wei, Y., Wilcox, D. T., Morand, T., Brahm, K., ... Massof, R. W. (2017). Outcomes of the veterans affairs low vision intervention trial ii (lovit ii): A randomized clinical trial. *JAMA Ophthalmology*, *135*(2), 96–104, https://doi.org/10.1001/jamaophthalmol.2016.4742.

Tang, W., Zhuang, S., & Liu, G. (2014). Comparison of visual function after multifocal and accommodative IOL implantation. *Eye Science*, *29*(2), 95–99.

Wandell, B. A. (2011). The neurobiological basis of seeing words. *Annals of the New York Academy of Sciences*, *1224*(1), 63–80, https://doi.org/10.1111/j.1749-6632.2010.05954.x.

Watson, A. B. (2017). QUEST+: A general multidimensional Bayesian adaptive psychometric method. *Journal of Vision*, *17*(3):10, 1–27, https://doi.org/10.1167/17.3.10. [PubMed] [Article]

Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in Cognitive Science*, *15*(4), 160–168, https://doi.org/10.1016/j.tics.2011.02.005.

Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, *63*(8), 1293–1313.

Yu, D. Y., Cheung, S. H., Legge, G. E., & Chung, S. T.

L. (2007). Effect of letter spacing on visual span and reading speed. *Journal of Vision*, *7*(2):2, 1–10, https://doi.org/10.1167/7.2.2. [PubMed] [Article]

Appendix A: Algorithm of the qReading method

In this section we describe the detailed algorithm of the qReading method. First, it characterizes the reading speed as a function of print size, with a three-parameter exponential function:

$$
\begin{aligned}
\log 10&(speed(size)) \\
&= \log10\left(\frac{60}{\tau(size)}\right) \\
&= \log10(\alpha) - (\log10(\alpha) - \log10(\alpha_C)) \\
&\quad \times \exp\left(-\frac{(\log10(size) - \log10(\kappa))}{\eta}\right), \quad \text{(A1)}
\end{aligned}
$$

Second, the qReading method defines the parameter space for all the possible reading curves $\theta = (\alpha, \kappa, \eta)$ and a prior distribution of parameters $p_0(\theta)$ representing the experimenter's prior knowledge of the probability of different reading curves. In the current study, we used a noninformative prior (flat prior over the entire parameter space).

Third, the method defines the stimulus space that contains all possible print sizes and presentation durations to be tested in the experiment $x = (size, duration)$ and a psychometric function in each stimulus condition, conditioned on the parameters of the reading curve:

$$
\begin{aligned}
p(r = &\text{correct}|x, \theta) \\
&= \Psi(x, \theta) = \gamma\lambda + (1 - \lambda) \\
\times &\left(\gamma + (1 - \gamma)\left(1 - \exp\left(-\left(\frac{duration}{\tau(size)}\right)^{\beta}\right)\right)\right), \quad \text{(A2)}
\end{aligned}
$$

where $\tau(size)$ is the threshold exposure duration corresponding to 80.3% correct and depends on the print size of the stimulus (Equations 1 and 2), γ and λ represent the guessing rate, lapse rate of the observer when performing the task, and β is slope of the psychometric function. The probability of an incorrect response is $p(r = \text{incorrect}|x, \theta) = 1 - \Psi(x, \theta)$. Combined together, Equations 1, 2, and A1 can model the response accuracy of the observer in any stimulus condition x. The equations also provide the conditioned joint probability of the parameter θ given the observed response to a stimulus x. In this way, the qReading test directly estimates the entire reading speed curve define by three parameters, instead of

measuring reading speed at one print size once a time, therefore greatly improving the testing efficiency.

Fourth, the qReading method calculates the mutual information[2] (Cover & Thomas, 1991; Kujala & Lukka, 2006) for each possible stimulus x,

$$I_{t-1}(\theta;r) = h\left(\int p_{t-1}(\theta)\Psi(x,\theta)d\theta\right) - \int p_{t-1}(\theta)h(\Psi(x,\theta))d\theta, \quad (A3)$$

where $p_{t-1}(\theta)$ with $t = 1, 2, 3\ldots$ is the prior knowledge about θ before the tth trial, and $h(p) = -p\log(p) - (1-p)\log(1-p)$ is the entropy of a distribution p. A one-step-ahead search determines the optimal stimulus x_t condition to be used in the next (tth) trial, by maximizing the expected information gain over the entire stimulus space:

$$x_t = \underset{x}{\mathrm{argmax}}(I_{t-1}(\theta;r)). \quad (A4)$$

Therefore, the qReading method avoids large regions of the stimulus space that provide little information to our current knowledge about θ.

Fifth, Bayesian update is used to accumulate the information collected in each trial throughout the measurement. After the response of the observer is collected in trial t, the posterior distribution of the reading function parameters $p_t(\theta)$ is computed, given the evidence provided by the observer's response $r =$ "correct" or "incorrect" to the stimulus $x_t = (size, duration)$ in the trial and the prior knowledge about parameters $p_{t-1}(\theta)$ via Bayes' rule:

$$p_t(\theta) = p_t(\theta|r_x) = \frac{p_{t-1}(\theta)p(r|x,\theta)}{\sum_\theta p_{t-1}(\theta)p(r|x,\theta)}. \quad (A5)$$

Lastly, the qReading method reiterates Steps 4 and 5 until it reaches a predefined number of trials. Then the estimated parameters θ as well as reading speed (Equation A1) are calculated based on the posterior distribution $p(\theta)$.

Appendix B: Procedure to eliminate the dependency of the estimated speeds across print size conditions

Because an exponential reading function with three parameters was used (Equation 1), the reading speeds obtained from the qReading method are not independent across print size conditions. To compute the correlation of the reading speeds estimated by the qReading and Psi methods, we carried out the following procedure for each observer to eliminate the dependency across conditions (Hou et al., 2016): (1) The reading speeds in the six print size conditions were first derived from the reading curves obtained with the qReading method; (2) for each of the six print sizes, randomly select one qReading run (out of eight, without replacement) and obtain the reading speed at that print size; (3) compute the correlation coefficient between the reading speeds in Step 2 with those obtained by the psi method; and (4) repeat Steps 2 to 3 five hundred times and calculate the average correlation coefficient. In this procedure, the reading speeds at different sizes were from entirely different qReading runs and were not constrained by the exponential model.