Assessing reading performance in the periphery with a Bayesian adaptive approach: The qReading method

Timothy G. Shepard College of Optometry, Ohio State University, Columbus, OH, USA

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

Peter J. Bex Psychology, Northeastern University, Boston, MA, USA

Luis A. Lesmes Adaptive Sensory Technology, San Diego, CA, USA

Zhong-Lin Lu Psychology, Ohio State University, Columbus, OH, USA

Deyue Yu College of Optometry, Ohio State University, Columbus, OH, USA

Reading is a crucial visual activity and a fundamental skill in daily life. Rapid Serial Visual Presentation (RSVP) is a text-presentation paradigm that has been extensively used in the laboratory to study basic characteristics of reading performance. However, measuring reading function (reading speed vs. print size) is time-consuming for RSVP reading using conventional testing procedures. In this study, we develop a novel method, qReading, utilizing the Bayesian adaptive testing framework to measure reading function in the periphery. We perform both a psychophysical experiment and computer simulations to validate the qReading method. In the experiment, words are presented using an RSVP paradigm at 10° in the lower visual field. The reading function obtained from the qReading method with 50 trials exhibits good agreement (i.e., high accuracy) with the reading function obtained from a conventional method (method of constant stimuli [MCS]) with 186 trials (mean root mean square error: 0.12 log10 units). Simulations further confirm that the qReading method provides an unbiased measure. The qReading procedure also demonstrates excellent precision (half width of 68.2% credible interval: 0.02 log10 units with 50 trials) compared to the MCS method (0.03 log10 units with 186 trials). This investigation establishes that the qReading method can adequately measure the reading function in the normal periphery with high accuracy, precision, and efficiency, and is a potentially valuable tool for both research and clinical assessments.

Introduction

Reading is an important visual activity and a fundamental skill that can be greatly affected by vision loss. Difficulty in reading has been reported as the main complaint of patients who attend low-vision clinics (Crossland, Gould, Helman, Feely, & Rubin, 2007), and improving reading ability is the main objective of the patients who seek low vision rehabilitation (Elliott et al., 1997; Owsley, McGwin, Lee, Wasserman, & Searcey, 2009). Reading performance is also positively correlated with the subjective quality of life for patients with central vision loss (Hazel, Petre, Armstrong, Benson, & Frost, 2000; Mitchell et al., 2008).

Although visual acuity is the most common functional vision endpoint, it can be insensitive to some retinopathies and their progression, especially in early disease stages (Klein, Wang, Klein, Moss, & Meuer, 1995; Sunness et al., 1997). Sometimes, when reading performance is impaired, visual acuity can be still within the normal limits or have little change (Crossland, Culham, & Rubin, 2004, 2005). In this regard, reading performance may be a better functional vision endpoint given its high sensitivity to both visual impairment (Cacho, Dickinson, Smith, & Harper,

Citation: Shepard, T. G., Hou, F., Bex, P. J., Lesmes, L. A., Lu, Z.-L., & Yu, D. (2019). Assessing reading performance in the periphery with a Bayesian adaptive approach: The qReading method. *Journal of Vision*, *19*(5):5, 1–14, https://doi.org/10.1167/19.5.5.

https://doi.org/10.1167/19.5.5 Received March 16, 2018; published May 6, 2019 ISSN 1534-7362

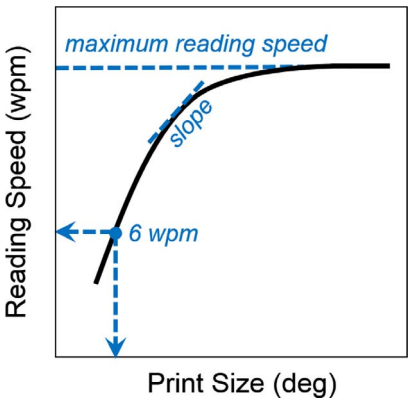

Figure 1. The reading curve shows reading speed (words per minute [wpm]) versus print size (lowercase x-height in degrees). In the present paper, the reading curve is described by an exponential function with three parameters: the asymptotic performance level (corresponding to maximum reading speed), the print size corresponding to a reading speed of six wpm, and the slope that controls the changing rate of the reading curve.

2010; Richter-Mueksch, Stur, Stifter, & Radner, 2006) and physical characteristics of text (Legge, 2007).

In both the clinics and laboratory, most reading tests aim to measure reading speed in a range of print sizes to construct a reading function (Figure 1). Legge and Bigelow (2011) have thoroughly reviewed the relationship between reading speed and print size. Each reading function contains a point corresponding to the critical print size below which reading speed begins to decline sharply and above which any further increase in print size does not lead to faster reading speed (Legge & Bigelow, 2011; Whittaker & Lovie-Kitchin, 1993). This relationship holds for normally-sighted people and patients with central vision loss (e.g., Chung, 2011; Chung, Mansfield, & Legge, 1998). Given this well-established relationship, identifying changes of the reading function relative to age-matched controls may help diagnose the nature of visual deficits in patients.

Several reading tests are available to clinicians to evaluate the impact of visual impairment on reading. Early reading assessments such as the Sloan Continuous Text Read Cards (sentence-based; Sloan & Brown 1963) and the Bailey–Lovie Near Reading Card (single word–based; Bailey & Lovie, 1980) allow clinicians to determine the magnification necessary to read normal text sizes. Later, the MNREAD test with a similar concept was introduced (Legge, Ross, Luebker, & Lamay, 1989). Other standardized tests include the Pepper Test (Baldasare, Watson, Whittaker, & Miller-Shaffer, 1986), Radner reading test (Radner et al., 1998), Jaeger reading test (Runge, 2000), and SKREAD (MacKeben, Nair, Walker, & Fletcher, 2015). The MNREAD test is among the most commonly administered reading tests in low-vision clinics.

Typically, in a reading test, patients are instructed to read printed sentences or groups of words while the clinician takes on the cumbersome task of covering/revealing a testing stimulus and recording both the reading time and accuracy. In this case, the performance measure (i.e., reading speed) is susceptible to "glitches" (Rubin, 2013). For example, if the clinician starts the stopwatch too soon before the patient is ready or if the patient repeats part of a sentence or makes corrections during reading, the estimated reading speed (calculated from the recorded reading duration) would not be accurate. The performance measure is also affected by the patient's articulation rate. Additionally, card-based tests require precisely calibrated and controlled viewing conditions (e.g., external lighting and viewing distance) to ensure the accuracy of the measurements. The number of tests that can be administered is determined by the number of unique versions of testing charts, which places an upper limit on the number of conditions that can be assessed (e.g., OS vs. OD vs. OU or high vs. low light level, etc.) and may frustrate efforts in assessing remediation or progression of disease over time.

Rapid Serial Visual Presentation (RSVP) is an alternative text presentation paradigm, primarily used in the laboratory. In an RSVP task, a rapid sequence of words is presented at a fixed location on a video display (Rubin, 2013). For both normal- and low-vision observers, RSVP reading speed is typically much higher than that for reading a page or a block of text (Rubin & Turano, 1992, 1994; Yu, Park, Gerold, & Legge, 2010), which can be at least partly attributed to the reduced need for eye movements in RSVP reading. In the laboratory, researchers use the RSVP paradigm to control where text is presented on the retina and provide precise control on word exposure duration. Although it is technically feasible to vary exposure duration word by word, the typical RSVP paradigm uses a fixed duration for word presentation throughout a sentence/trial, making it easier to compute reading speed. In comparison to page reading, RSVP reading eliminates and, therefore, does not assess eye movements, natural rhythm of eye fixations, and trans-saccadic integration. However, the RSVP paradigm offers precise control of word exposure duration, making the reading speed measurement impervious to many uncontrollable factors. For instance, observers can initiate a trial whenever they are ready. The exposure duration of words in each trial is precisely controlled, which is recorded as reading time. Observers can take time to report the words and make corrections, which does not affect the measurement of their reading speed.

With conventional testing procedures (e.g., non-adaptive methods, such as the method of constant

stimuli (MCS; Chung et al., 1998; Legge et al., 2007) or adaptive staircase methods (Fine, Peli, & Reeves, 1997; Pelli & Tillman, 2007), obtaining a reading function (reading speed vs. print size) for RSVP reading requires measuring reading performance repeatedly over a range of predetermined stimulus conditions, which can be overly time-consuming for clinical use. For instance, the MCS procedure typically takes about 1 hr to capture the reading curve across a range of print sizes in normal vision. The testing time is even greater for low-vision patients, who may take more time to read and fatigue more easily. Hence, there is a clear need for developing efficient testing methods aiming at easy administration and short testing duration while providing accurate and precise estimate of the reading function.

Here we adopt a Bayesian adaptive testing framework (Lu & Dosher, 2013) to develop an efficient testing method, qReading, to measure the reading function in the periphery for RSVP reading. The Bayesian adaptive procedure was first applied in the landmark development of the QUEST method (Watson & Pelli, 1983) and is now extensively used in psychophysics (e.g., Alcalá-Quintana et al., 2007; García-Pérez & Alcalá-Quintana, 2007; Hou, Lesmes, Bex, Dor, & Lu, 2015; King-Smith, Grigsby, Vingrys, Benes, & Supowit, 1994; King-Smith & Rose, 1997; Lesmes, Jeon, Lu, & Dosher, 2006; Lesmes, Lu, Baek, & Albright, 2010; Remus & Collins, 2007, 2008; Snoeren & Puts, 1997; Watson, 2017). To provide efficient assessment while maintaining high precision and accuracy of the measurements, the qReading method combines Bayesian inference and an observer model to capture the regularities in reading data and adopts the information-theoretic framework to select the most informative testing stimuli. In this study, the reading function, reading speed (converted from exposure duration) versus print size, is described by an exponential function with three parameters. With the information collected from each trial (i.e., testing stimulus and the corresponding response), the estimation on an individual's reading function is refined on a trial-by-trial basis. The qReading method updates the joint posterior distribution of the parameters of the reading function and, in turn, updates the estimated reading function. Subsequently, a one-step-ahead search is performed to select the print size and exposure duration that optimizes the information gain for the upcoming trial.

In this report, we describe a psychophysical experiment and computer simulations to validate the qReading method when measuring the reading function for RSVP reading in the periphery. We demonstrate that the qReading method can adequately measure the reading function in the periphery with high accuracy, precision, and efficiency and could be a useful tool for the assessment of visual function of both normal- and low-vision patients.

General methods

Observers

Five native English-speaking adults aged 19–25 years with normal or corrected-to-normal vision were recruited. None of the observers reported difficulties in reading and were all naive to the purpose of the experiment. The Ohio State University's institutional review board approved the research protocol, and the procedures complied with the Declaration of Helsinki. The experiments were undertaken with written consent of each observer.

Apparatus and stimuli

The experiments were programmed and controlled using MATLAB (MathWorks, Ltd.) and PsychToolbox-3 (Brainard, 1997; Kleiner, Brainard, Pelli, Ingling, Murray, & Broussard, 2007; Pelli, 1997) on a MacBook Pro. The display was a ViewSonic Graphics Series G220fb CRT monitor (size of the screen, 40.5 × 30.2 cm) with a 1,280 × 1,024-pixel resolution and 85-Hz refresh rate. An EyeLink 1000 eye tracker (SR Research, Ltd., Ottawa, Ontario, Canada) was used to monitor the eye position of each observer. All testing was conducted binocularly in a dark room at a viewing distance of 60 cm, which was preserved with the use of a chinrest. Observers were asked to maintain stable fixation on a fixation line that ran horizontally across the center of the display for the entire trial. Words were presented at 10° below the fixation line. All characters were lowercase English letters and depicted as black characters on a white background with a luminance of 115 cd/m^2. A monospaced (i.e., fixed width) Courier font and the standard letter spacing (1.16 × x-width) were used.

RSVP reading task

The RSVP task (e.g., Yu, Cheung, Legge, & Chung, 2007) was used as the reading task when comparing a conventional testing method (MCS method; see section "The MCS method") and the qReading method. On a given trial, a sentence was randomly selected without replacement from the sentence pool that consisted of 1,180 sentences, each containing 10 words. Each word had six letters or fewer in length. The text is simple enough to ensure that the

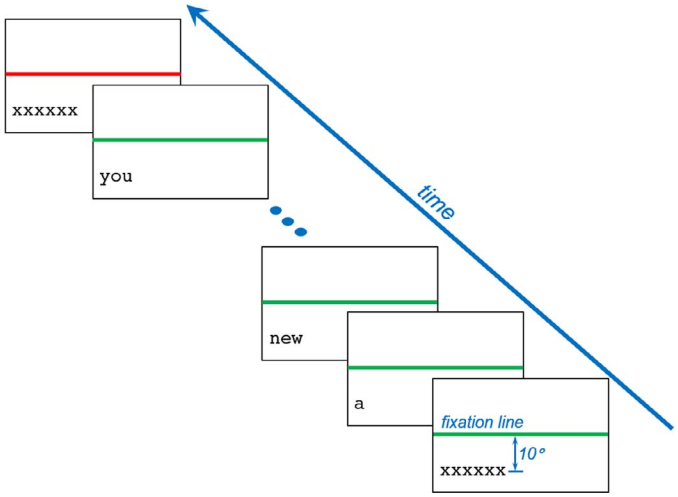

Figure 2. A schematic diagram of the RSVP task.

performance is not limited by vocabulary or syntax, and no punctuation was presented in any of the sentences. Words of the sentence were presented serially (i.e., one word at a time) at the same location (left justified) on the display with a specified exposure duration. A mask, "xxxxxx," was presented before the first word of the sentence to indicate the print size and location of the stimuli, and it also appeared after the last word of each sentence (see Figure 2). Observers were instructed to read the words aloud while maintaining stable fixation along the horizontal fixation line. Eye movements were monitored by the EyeLink 1000 eye tracker sampling at 1,000 Hz. Although viewing was binocular, only the left eye was tracked for each observer. Horizontal eye movements along the fixation line were permitted. The trial was cancelled and repeated with a different sentence if fixation deviated (±1°) above or below the fixation line. Deviation of the fixation was detected in 7.7% of the qReading trials and 5.0% of the other reading trials, averaged across observers.

The MCS method

When measuring the reading function for RSVP reading, the conventional testing procedure typically measures reading performance repeatedly for a range of predetermined print sizes. In this study, MCS was used to obtain the conventional baseline because it provides an estimate of the entire psychometric function (reading accuracy as a function of exposure duration) at each print size. Five print sizes, 0.60°, 0.98°, 1.58°, 2.58°, and 4.20°, were tested. At each print size, six exposure durations, 82, 141, 247, 435, 765, and 1,329 ms/word, were used. For the smallest print size, a longer duration of 2,318 ms was added to ensure that the observers were able to attain 80% or

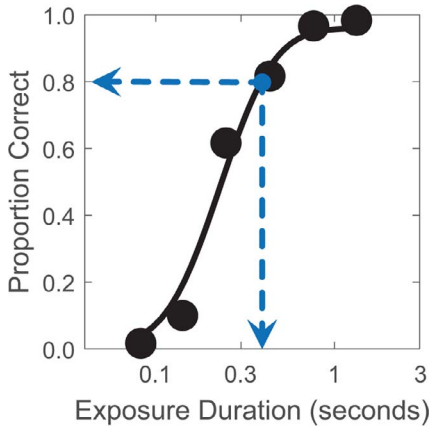

Figure 3. Example plot of reading accuracy (proportion of words read correctly) as a function of exposure duration. The data are collected from one of our observers at one print size using the MCS method and fitted with a one-parameter cumulative Gaussian function. The arrows indicate the 80% accuracy (threshold) and the threshold exposure duration.

higher reading accuracy at the longest duration. Therefore, there were 31 stimulus conditions (5 print sizes × 6 exposure durations + 1 additional exposure duration at the smallest print size). There were six blocks of trials. Each block had 31 trials with one trial per condition. The order of stimulus conditions was randomized in each block. Collectively, the six blocks contained 186 trials and, with breaks, took about 1 hr. The data collected at each print size were fitted with a cumulative Gaussian function (Equation 1), where the guess rate ($\gamma = 0.01$), slope ($\beta = 0.24$), and lapse rate ($\lambda = 0.04$) were fixed[1] and the shift (τ) was free to vary:

$$\Psi(\log_{10} duration) = \gamma\lambda + (1-\lambda)\left(\gamma + (1-\gamma)\Phi\left(\frac{\log_{10} duration - \tau}{\beta}\right)\right), \quad (1)$$

where Φ is the cumulative distribution function of the standard Gaussian distribution. The same equation was also used for psychometric functions in the qReading method. Figure 3 illustrates an example data set collected at one print size from one observer using the MCS method along with the fitted psychometric function. Reading accuracy was plotted as a function of exposure duration. Reading speed was computed based on the threshold exposure duration at which observer's response is 80% correct according to the fitted curve: reading speed (wpm) = 60/(threshold exposure duration in seconds). The reading speed versus print size function (see Equation 2) was then estimated based on the reading speeds measured at the five print sizes.

The qReading method

Similar to a parallel study (Hou et al., 2018), the qReading method in the present study consists of six components. The first component is the functional form of the reading curve. We use an exponential function to model the reading curve (Equation 2). Similar nonlinear functions have also been used by others (e.g., Bernard, Scherlen, & Castet, 2007; Cheung, Kallie, Legge, & Cheong, 2008; Seiple, Szlyk, Memahon, Pulido, & Fishman, 2005; Yu et al., 2010). The function has three parameters: asymptotic performance level in terms of threshold exposure duration (AD, corresponding to the maximum reading speed), the print size at which the reading speed is six wpm (CS), and slope (DC, the slope of the function that determines the changing rate of the reading curve):

$$\log_{10} ReadingSpeed = \log_{10}\left(\frac{60}{AD}\right) - \log_{10}\left(\frac{10}{AD}\right) e^{-\left(\frac{\log_{10} PrintSize - \log_{10} CS}{DC}\right)}. \quad (2)$$

Second, a three-dimensional parameter space, $\vec{\theta} = (AD, CS, DC)$, is defined to encompass all potentially observable reading curves for the target population and testing condition (10° eccentricity in the normal periphery for the present study). The joint prior distribution of the three parameters is determined based on a priori knowledge of the probability of different reading curves. Specifically, the ranges of the three parameters are −1.12 to 0.18 (corresponding to 800 to 40 wpm)^2 for $\log_{10}AD$, −1 to 0 (corresponding to 0.1° to 1° print size) for $\log_{10}CS$, and −1.5 to 0 for $\log_{10}DC$. In order to perform all the computations in real time, for each parameter, 25 values were evenly sampled (in log10 units) within its range. A finer grid in the parameter space (100 values evenly sampled over each parameter's range) was used in the data analysis.

Third, a two-dimensional stimulus space is constructed to span broad ranges of print size and exposure duration. There are 25 log-spaced print sizes between 0.3° and 4.2° and 42 log-spaced exposure durations between 24 and 3,000 ms (2–255 frames) in the stimulus space. For each print size, a psychometric function (see Equation 1) is defined based on the parameters of the reading curve. Consistent with the MCS method, the threshold exposure duration at 80% correct was used to derive reading speed from the psychometric functions.

Fourth, a one-step-ahead search is performed to select the stimulus condition (a combination of print size and exposure duration) for the subsequent trial with the goal of optimizing the expected information gain on the reading curve (Baek, Lesmes, & Lu, 2016; CoboLewis, 1996; Kontsevich & Tyler, 1999; Kujala & Lukka, 2006; Lesmes et al., 2006).

Fifth, following the observer's response to the stimulus in each trial, the joint posterior distribution of the three parameters is updated using Bayes' rule. The posterior distribution then serves as the new prior for the upcoming trial.

Last, the qReading method repeats the fourth and fifth steps until a certain criteria is met to achieve an accurate and precise estimation of the reading curve.

Because each sentence contains 10 words and, for each word, the observer generates one response, which could be either correct or incorrect (including missing the word), there are 10 responses from each trial. Therefore, the joint posterior distribution of the three parameters is actually updated 10 times following each trial. There are 50 trials in the qReading block with an average testing duration of approximately 12 min.

Experimental design

The experiment consisted of one session with seven blocks. The qReading block was placed in the middle of the session (i.e., surrounded by MCS blocks) except for observer S2 who did the qReading block after the second MCS block. Twenty practice trials were given to each observer in the beginning of the session, which were not included in the analysis.

Psychophysical validation

Figure 4 shows the reading curves obtained from the qReading and the MCS methods for the five observers. The columns show how the estimated reading curve evolves in 10-trial increments throughout the block. As shown in Figure 4, the reading curves obtained with the two methods are comparable. To quantify the similarity of the two reading curves, we calculated the root mean square error (RMSE) of the reading speeds (in log10 units) obtained with the 50 qReading trials and the 186 MCS trials, collapsed across all observers ($m = 5$) for all print sizes ($n = 5$)[3] that were measured in both methods (i.e., five print sizes between 0.6° and 4.2°):

$$RMSE = \sqrt{\frac{\sum_{j=1}^{m}\sum_{i=1}^{n_j}\left(ReadingSpeed\_qReading_{i,j} - ReadingSpeed\_MCS_{i,j}\right)^2}{\sum_{j=1}^{m} n_j}}. \quad (3)$$

We found that the RMSE was 0.16, 0.16, 0.13, 0.13, and 0.12 log10 units after 10, 20, 30, 40, and 50 trials with the qReading procedure, respectively.

We calculated the area under the curve (AUC) for the estimated reading curves from both methods. Specifically, we used log10 scale for both the x- and y-axes (i.e., reading speed in log10(wpm) and print size in

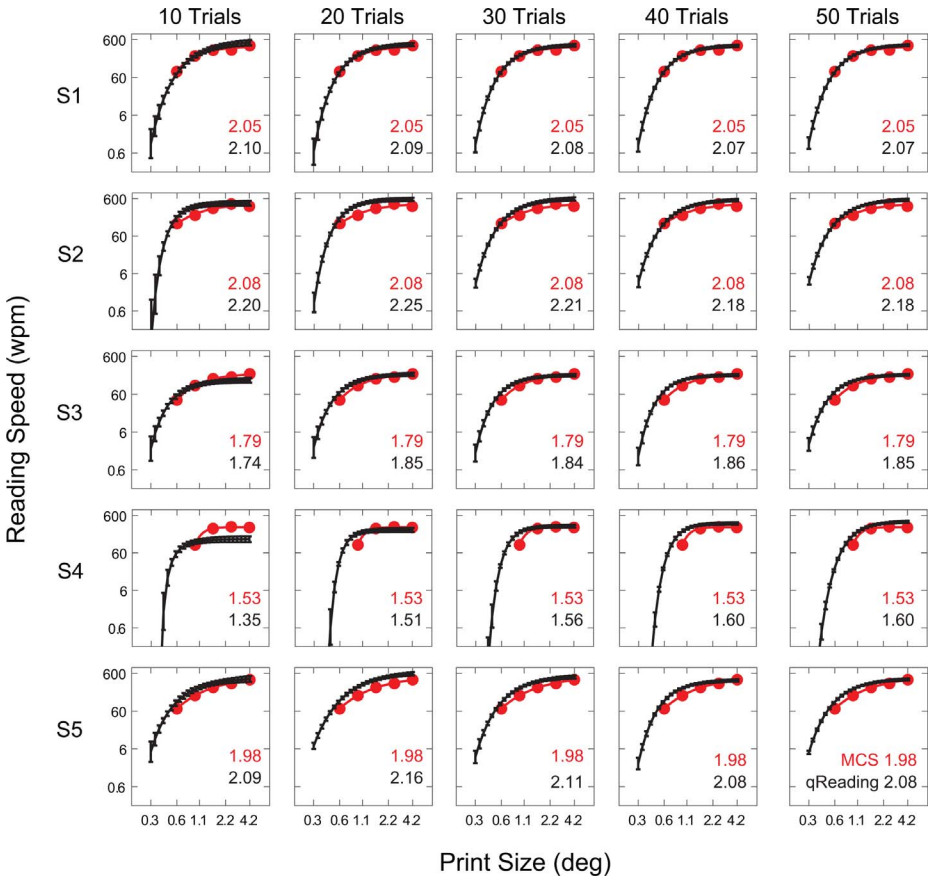

Figure 4. Reading curves from the qReading and the MCS methods. Each row presents reading data from one observer. The columns show how the qReading curve evolves in 10-trial increments throughout the block. The black curves are the estimated reading curves using the qReading method. The error bars represent the ±68.2% HWCIs. The red filled circles represent the estimated reading speeds from the MCS method with 186 trials. In each panel, we list the AUC values from the qReading (black text) and the MCS methods (red text).

log10(°)), and calculated the total area enclosed by the reading curve and the horizontal line through log10(1 wpm) in the print size range assessed by both methods (i.e., 0.6° to 4.2°). The individual data with the AUC calculation (in the square of log10 units) are shown in Figure 4.

A Bland–Altman plot (Bland & Altman, 1986, 1999) was used to evaluate the agreement between the estimates from the MCS and the qReading methods. Figure 5 plots the AUC difference between the reading curves measured with the MCS and the qReading methods versus the mean AUC from the two methods. We estimated an agreement interval in which 95% of the differences between the AUCs measured with the two methods reside. Across all five observers, the difference between the AUCs from the two methods always fell within the limits of agreement (–0.14 and –0.01 square log10 units). Figure 5 also shows that the AUCs from the qReading method were always slightly larger (0.07 square log10 units, 4%) than those from the MCS method, indicating a small but consistent bias. After subtracting the bias from the qReading measure, we found a very good agreement between the two methods as indicated by the small limits of agreement. As shown later (see the "Simulation" section), we further evaluated the potential bias of the qReading

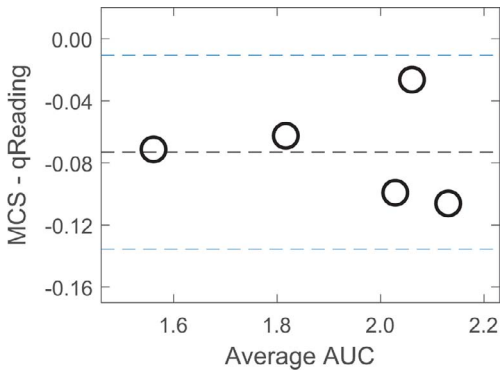

Figure 5. A Bland–Altman plot. The *x*-axis denotes the mean of the AUCs measured with the MCS and the qReading methods. The *y*-axis indicates the difference between the corresponding AUCs. The blue dashed lines represent the 95% limits of agreement. The black dashed line indicates bias.

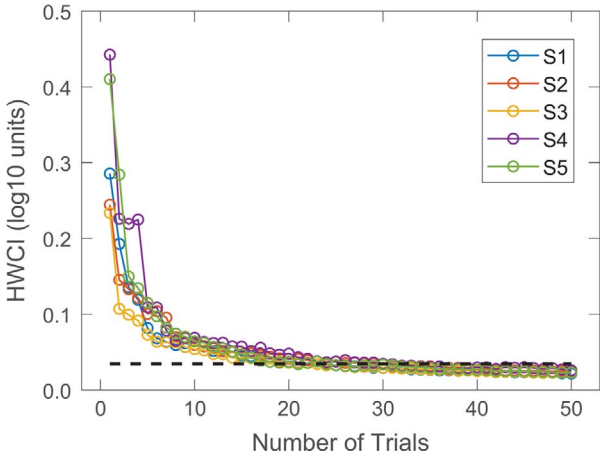

Figure 6. The average 68.2% HWCI of the estimated reading function from the qReading method plotted as a function of the number of trials for five observers. The black dashed line denotes the 68.2% HWCI (group average) of the estimated reading function from 186 trials with the MCS method. The HWCIs shown here are calculated across the print sizes assessed in both the qReading and the MCS methods.

method in simulations and found that the qReading method was essentially an unbiased procedure. It indicates that the small bias shown here may reflect improved performance during the test. For instance, the observers' reading speeds often are slower in the first one or two MCS blocks compared to the rest of the session (including the qReading block) due to learning.

The half width of the credible interval (HWCI) of the posterior distribution of the estimated performance can provide a measure of precision in a single run of a procedure. It is a proper measure of precision especially for the present study because we only obtained one measurement (reading function) from each observer using each testing method. The 68.2% credible interval represents the interval in which the true value of the variable of interest lies with a 68.2% probability. The 68.2% HWCI is calculated as

$$\mathrm{HWCI}_{0.682} = \frac{\Phi^{-1}(0.841) - \Phi^{-1}(0.159)}{2}, \quad (4)$$

where Φ^{-1} represents the inverse cumulative distribution function of the posterior. We first calculated the HWCI at each print size and then computed the average across print sizes. Since 25 different print sizes were used during the reading curve measurement with the qReading method, the average HWCI for qReading was first calculated across the 25 print sizes for each individual. Overall, the HWCI decreased (i.e., measurement becomes more precise) as more qReading trials were added. With 10, 20, 30, 40, and 50 qReading trials, the average HWCI across observers was 0.15, 0.23, 0.11, 0.08, and 0.05 log10 units, respectively. When we evaluated only the five print sizes that were assessed in both the qReading and the MCS methods, the corresponding HWCIs were 0.06, 0.04, 0.03, 0.03, and 0.02 log10 units, respectively. Figure 6 shows the average HWCI across the common print sizes plotted as a function of the number of qReading trials for the five observers. The HWCI for the MCS method were generated with a bootstrapping technique. In each iteration, six sentences were sampled for each exposure duration with replacement at each print size (a total of 186 sentences). The data were then fit with the cumulative Gaussian function (Equation 1) to get an estimate of reading speed for each print size. After 3,000 iterations, the HWCI was calculated for each of the five print sizes. The average HWCI across print sizes and observers was 0.03 log10 units for the MCS method.

	$\log_{10}$AD	$\log_{10}$CS	$\log_{10}$DC
Sim1	−0.50	−0.57	−0.71
Sim2	−0.70	−0.40	−0.35
Sim3	−0.92	−0.48	−0.55

Table 1. Reading curve parameters for three simulated observers.

Simulations

We performed simulations to further evaluate the qReading procedure. These simulations allow us to test the basic assumptions of the qReading method (e.g., priors, parameter space for reading function, parameters of psychometric functions, etc.) and evaluate the performance of the algorithm (e.g., convergence of the posterior). Most importantly, the simulations allow us to evaluate whether the method is biased because we know the exact "truth" in the simulations but we don't know the "truth" in most experiments. We can generate numerous trials and blocks (e.g., a block containing 100 trials or 1,000 blocks with 50 trials per block) and compare each estimated reading curve to the ground truth.

To perform simulations, we first defined three simulated observers by assigning values to the three parameters (AD, CS, and DC) of their reading curves. The values were selected to cover a wide range of possible reading curves when considering peripheral reading performance in normally-sighted observers (see Table 1). The reading curves constructed by the three parameters are considered as the true reading curves which were used to generate the simulated observer's responses and acted as a yardstick to evaluate the accuracy of the qReading algorithm. All of the other parameter settings and procedure details (e.g., range of print size, range of exposure duration, strategy for

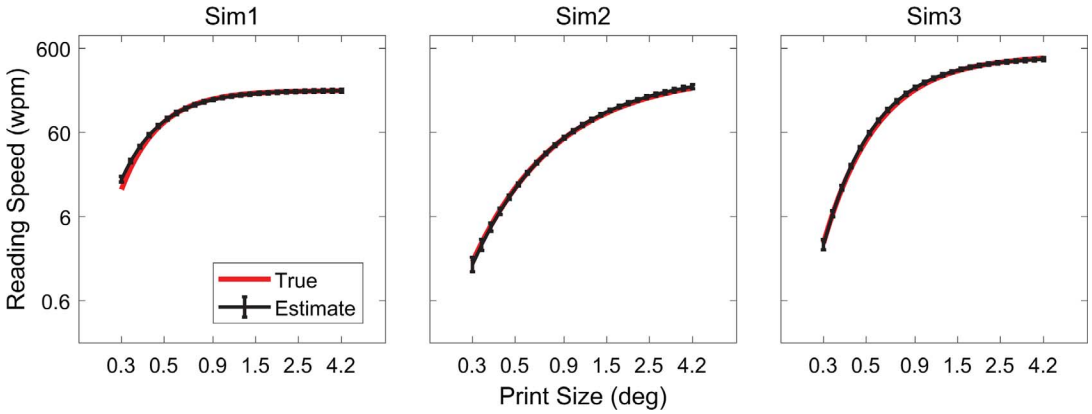

Figure 7. Simulated qReading performance in 100 trials for each of the three simulated observers. The black curve is the estimated reading function based on the simulated responses. The error bars represent ±68.2% HWCIs. The red curve represents the ground truth.

selecting stimulus condition, the prior distribution of the parameters) were the same as those in the psychophysical experiment.

In the first simulation, a 100-trial qReading block is generated for each simulated observer. Figure 7 shows the estimated and true reading functions. To quantify the similarity of the estimated reading function and the ground truth, we calculated the RMSE across all three simulated observers ($m = 3$) and 25 print sizes ($n = 25$). After 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100 trials, the RMSE was 0.05, 0.04, 0.04, 0.03, 0.03, 0.03, 0.03, 0.02, 0.02, and 0.02 log10 units, respectively.

Table 2 lists the estimated and true AUCs in the print size range of 0.3° to 4.2°. It shows how the AUC changes in increments of 10 trials throughout the entire block.

Figure 8 shows the average HWCI as a function of the number of trials for the three simulated observers. Similar to what we found in the human observers, the average HWCI decreased (i.e., qReading estimate becomes more precise) as more trials were added. Similar to the observations in the psychophysical experiment, the average HWCIs across the three simulated observers at trial 10, 20, 30, 40, and 50 were small: 0.07, 0.05, 0.04, 0.03, and 0.03 log10 units, respectively. The precision of the qReading method (HWCI) did not change significantly after 50 trials. The HWCIs at trial 60, 70, 80, 90, and 100 were 0.03, 0.02, 0.02, 0.02, and 0.02 log10 units, respectively. These simulations suggest that additional testing after 50 trials may not lead to significant improvement in accuracy and precision.

We also performed a separate simulation to further evaluate the accuracy of the estimated reading curve. In this simulation, each simulated observer was tested with the qReading method for 1,000 blocks with 50 trials per block. Estimation bias was then calculated. First, the AUC was calculated across the print size range of 0.3° to 4.2° after each trial in each block. Second, for each number of completed trials, the average AUC was computed across the 1,000 blocks. The absolute difference between the average AUC and the true AUC (i.e., the AUC calculated based on the true reading curve) was obtained and referred to as bias. Figure 9 plots the bias as a function of the number of trials for the three simulated observers. As the number of trials increased, the bias always dropped below 0.01 log10 units. This shows that the qReading method is essentially an unbiased procedure.

Discussion

In the current study, we present a Bayesian adaptive method, qReading, to measure the reading function in the peripheral visual field. The qReading utilizes a Bayesian adaptive framework to select the stimulus condition that optimizes the expected information gain prior to each trial, thus allowing for efficient assessment

Number of trials	10	20	30	40	50	60	70	80	90	100	Ground truth
Sim1	2.40	2.40	2.42	2.42	2.41	2.40	2.40	2.40	2.40	2.39	2.38
Sim2	1.91	1.94	1.93	1.93	1.93	1.92	1.92	1.92	1.92	1.92	1.92
Sim3	2.46	2.53	2.50	2.50	2.50	2.49	2.49	2.48	2.48	2.49	2.47

Table 2. AUCs (in the square of log10 units) calculated across print sizes (0.3° to 4.2°) based on one 100-trial qReading block for the three simulated observers.

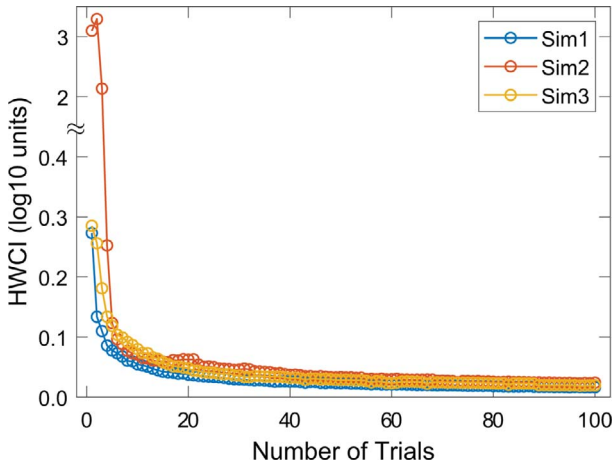

Figure 8. The average 68.2% HWCI as a function of the number of trials for the three simulated observers.

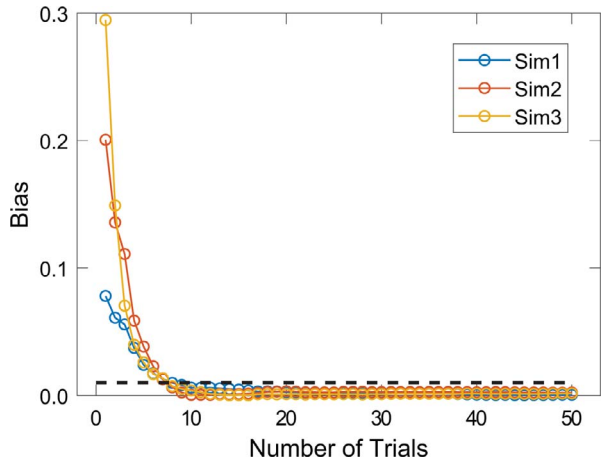

Figure 9. Estimation bias as a function of the number of trials for the three simulated observers (average across 1,000 blocks). The black dashed line denotes 0.01 log10 units.

of the reading function. Both psychophysical experiment and computer simulations were performed to validate the qReading method. Our investigation demonstrated that the qReading method can adequately measure the reading function in the periphery with high accuracy, precision, and efficiency. To assess accuracy, we compared the estimated reading functions measured using the qReading method to those obtained with the MCS method. We found that the qReading method is highly accurate with the RMSE dropping to 0.13 log10 units in about 30 trials. The high accuracy was further verified by computer simulations—after about 30 trials, the RMSE dropped below 0.04 log10 units. Our simulations also showed that the bias always drops below 0.01 log10 units after only eight trials, and after 30 trials, the qReading method introduced nearly zero bias in estimating the AUC of the reading function. We computed the HWCIs from both methods to evaluate their precision and found that it only requires, on average, about 30 qReading trials to reach the same precision achieved by 186 MCS trials. Simulations further confirmed the high precision of the qReading method—the average HWCIs fell below 0.05 log10 units in less than 30 trials.

Although, in the present study, 50 qReading trials were used to assess the accuracy and precision of the qReading method, our results suggested that fewer trials may be adequate to obtain an accurate and precise measurement of the reading function. For instance, with about 30 qReading trials (7 min), we can get an accurate estimate of the reading function and match the precision achieved by the 1-hr MCS test. One way to decide on the number of required trials and improve the efficiency of the test is to employ a stopping rule and conclude the testing when the stopping rule is satisfied. For example, the stopping rule can be based on HWCI. Once a predefined level of precision has been reached, we can end the testing.

Another factor that we need to consider in terms of further improving testing efficiency is the prior (joint prior distribution of the three reading-function parameters). In the present study, the qReading method adopted a broad prior that was determined based on some a priori knowledge of the probability of different reading functions for the target population and testing condition. Potentially, more informative priors can be used to further optimize the qReading method and improve its testing efficiency. Here, the reading curve is expressed as an exponential function. It is possible that the reading curve can be described with some other parametric model (e.g., hinged lines, Gaussian, and polynomial), a nonparametric model, or a combination thereof for different patient populations and conditions. As we collect more data with the qReading method in different populations, we will build a population-level structure to better tailor the prior and the model based on the demographics of each individual patient (e.g., age, eye disease history, education level, and level of language proficiency). Specifically, we can implement a hierarchical Bayesian framework that not only considers the responses from earlier trials in the current block but also incorporates responses collected from other individuals in the same population (Gu et al., 2016; Kim, Pitt, Lu, Steyvers, & Myung, 2014). We can also utilize existing information from the observer (such as visual acuity and practice trials) to further customize the prior.

One advantage of the qReading method is that the method can be easily modified to test reading in different conditions. In the current study, we measured the reading function in a specific reading condition: viewing simple sentences (10 common words per sentence, $\leq$6 letters per word) presented with the RSVP paradigm at 10° eccentricity in the periphery. In future studies, we can evaluate the reading function using

other variations of reading materials, such as words with a fixed length, words with various frequencies, sentences with scrambled word order, etc. We can also adopt other text-presentation methods (e.g., page reading); position text at various retinal locations; present text in different contrasts, spacings, fonts, colors, and directions; and employ different reading tasks. The easy modification of the qReading method will allow us to test these aforementioned alternative conditions. In a parallel study, Hou et al. (2018) developed and applied the qReading method to a word/nonword judgment task. After viewing a five-letter string, the observers were instructed to indicate if the letter string was a word or nonword. Although it can be considered as a truncated, one-word, silent version of the RSVP task, the lexical decision task used in the Hou et al. (2018) study only examines word recognition and ignores many nonvisual linguistic factors, such as grammar and context in reading. Similarly, they found that the qReading method was accurate, precise, and efficient compared to the corresponding conventional testing method. Arango et al. (2017) also applied the qReading method to another reading task in which participants indicated whether four-word sentences were logically true or false (Crossland, Legge, & Dakin, 2008). Both of these studies were performed in central vision where reading performance is much less limited by sensory factors (decreasing resolution, mislocations, and crowding) in comparison to peripheral vision (He, Legge, & Yu, 2013; Yu, Legge, Wagoner, & Chung, 2014). These studies demonstrate the flexibility in applying the qReading technique to efficient assessment of reading performance.

In everyday page reading, eye fixations are distributed across words in a nonhomogeneous fashion, especially for people with central vision loss (Calabrèse, Bernard, Faure, Hoffart, & Castet, 2016). Clusters of eye fixations form around words that are challenging to identify, reflecting the difficulty of trans-saccadic integration for these words. It has been shown that fixation clustering is a significant predictor of reading speed (Calabrèse et al., 2016). For RSVP reading, oculomotor measures were used only to monitor fixation stability. In other implementations of the qReading method that adopt, for example, page reading as the text presentation method, we will incorporate fixation clustering and other oculomotor measures in assessing reading performance. Doing so would allow us to characterize the reading performance in a more comprehensive way.

There are 246 million people worldwide with low vision (World Health Organization, 2014). A comprehensive, accurate, and individual-based reading assessment is crucial in developing effective rehabilitation methods and prescribing adaptive devices to assist these patients with their reading needs and improve their quality of life. Such assessment is also essential for monitoring the level of vision impairment throughout the course of the disease. With the conventional MCS testing method, administering comprehensive reading assessment and monitoring the change of reading performance as disease progresses in patients would present a great clinical burden. Here we address the issue by developing the qReading method. Collectively, the results of the psychophysical experiment and computer simulations demonstrate that the qReading method can adequately measure RSVP reading function in the periphery with high precision, accuracy, and efficiency.

Keywords: peripheral reading, low vision, adaptive methods

Acknowledgments

This research was supported by National Institutes of Health grants EY025658 (DY) and EY021553 (ZLL) and the National Natural Science Foundation of China (NSFC81600764 to FH). The authors gratefully acknowledge Carter Bensman's help on subject recruitment and data collection.

Commercial relationships: HF, PB, LAL, ZLL, and DY own intellectual property rights on qReading technology. PB, LAL, and ZLL have equity interest in Adaptive Sensory Technology, Inc. LAL holds employment at AST.
Corresponding author: Deyue Yu.
Email: deyueyu@gmail.com.
Address: College of Optometry, Ohio State University, Columbus, OH, USA.

Footnotes

[1] In the present study, we made the assumption that the shape of the underlying psychometric function for reading is known and constant across print sizes. First, we derived reading speeds by fitting the RSVP data using the cumulative Gaussian function that has either one free parameter (i.e., shift) or four free parameters (i.e., guess rate, slope, lapse rate, and shift). We found that fits of the two models were not significantly different (AUC; see "Psychophysical validation" section for calculation details; $p = 0.36$). Therefore, the one-parameter fitting routine was used in the present study. Second, the values of the three fixed parameters were determined based on previous empirical data collected in a similar testing condition for the same target population. In addition, we performed simula-

tions to confirm that adopting different values of slope, lapse rate, and guess rate (over large ranges) had little effect on our findings for the target population.

[2] For observer S1, we used the range of −1 to 0.30 (corresponding to 600 to 30 wpm) for parameter $\log_{10}AD$ during the data collection. To be consistent with the other observers, the range of −1.12 to 0.18 was used for data analysis.

[3] Due to the poor performance of the observer, we were not able to obtain a reliable estimation of S4's reading speed at the smallest print size (0.60°) using the MCS method. Therefore, n is equal to four for S4. The same is true for AUC and HWCI calculations.

References

Alcalá–Quintana, R., & García-Pérez, M. A. (2007). A comparison of fixed-step-size and Bayesian staircases for sensory threshold estimation. *Spatial Vision*, *20*(3), 197–218.

Arango, T., Hou, F., Lesmes, L., Yu, D., Lin, 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]

Baek, J., Lesmes, L. A., & Lu, Z. L. (2016). qPR: An adaptive partial-report procedure based on Bayesian inference. *Journal of Vision*, *16*(10):25, 1–23, https://doi.org/10.1167/16.10.25. [PubMed] [Article]

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.

Baldasare, J., Watson, G. R., Whittaker, S. G., & Miller-Shaffer, H. (1986). The development and evaluation of a reading test for low vision individuals with macular loss. *Journal of Visual Impairment & Blindness*, *80*(6), 785–789.

Bernard, J. B., Scherlen, A.-C., & Castet, E. (2007). Page mode reading with simulated scotomas: A modest effect of interline spacing on reading speed. *Vision Research*, *47*(28), 3447–3459.

Bland, J. M., & Altman, D. G. (1986). Statistical method for assessing agreement between two methods of clinical measurement. *The Lancet*, *2*, 307–310.

Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, *8*, 135–160.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436, https://doi.org/10.1163/156856897x00357.

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., Bernard, J. B., Faure, G., Hoffart, L., & Castet, E. (2016). Clustering of eye fixations: A new oculomotor determinant of reading speed in maculopathy. *Investigative Ophthalmology & Visual Science*, *57*(7), 3192–3202.

Cheung, S.-H., Kallie, C. S., Legge, G. E., & Cheong, A. (2008). Nonlinear mixed-effects modeling of MNREAD data. *Investigative Ophthalmology & Visual Science*, *49*, 828–835.

Chung, S. T. L. (2011). Improving reading speed for people with central vision loss through perceptual learning. *Investigative Ophthalmology & Visual Science*, *52*(2), 1164–1170, https://doi.org/10.1167/iovs.10-6034.

Chung, S. T. L., 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, https://doi.org/10.1016/S0042-6989(98)00072-8.

CoboLewis, A. B. (1996). An adaptive method for estimating multiple parameters of a psychometric function. *Journal of Mathematical Psychology*, *40*(4), 353–354.

Crossland, M. D., Culham, L. E., & Rubin, G. S. (2004). Fixation stability and reading speed in patients with newly developed macular disease. *Ophthalmic and Physiological Optics*, *24*(4), 327–333, https://doi.org/10.1111/j.1475-1313.2004.00213.x.

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

Crossland, M. D., Gould, E. S., Helman, C. G., Feely, M. P., & Rubin, G. S. (2007). Expectations and perceived benefits of a hospital-based low vision clinic: Results of an exploratory, qualitative research study. *Visual Impairment Research*, *9*(2–3), 59–66, https://doi.org/10.1080/13882350701643582.

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*, 1–9, https://doi.org/10.1186/1744-9081-4-14.

Elliott, D. B., TrukoloIlic, M., Strong, J. G., Pace, R.,

Plotkin, A., & Bevers, P. (1997). Demographic characteristics of the vision-disabled elderly. *Investigative Ophthalmology & Visual Science*, *38*(12), 2566–2575.

Fine, E. M., Peli, E., & Reeves, A. (1997). Simulated cataract does not reduce the benefit of RSVP. *Vision Research*, *37*(18), 2639–2647.

García-Pírez, M. A., & Alcalá-Quintana, R. (2007). Bayesian adaptive estimation of arbitrary points on a psychometric function. *British Journal of Mathematical & Statistical Psychology*, *60*, 147–174, https://doi.org/10.1348/000711006x104596.

Gu, H. R., Kim, W., Hou, F., Lesmes, L. A., Pitt, M. A., Lu, Z. L., & Myung, J. I. (2016). A hierarchical Bayesian approach to adaptive vision testing: A case study with the contrast sensitivity function. *Journal of Vision*, *16*(6):15, 1–17, https://doi.org/10.1167/16.6.15. [PubMed] [Article]

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 & Visual Science*, *41*(6), 1309–1315.

He, Y., Legge, G. E., & Yu, D. (2013). Sensory and cognitive influences on the training-related improvement of reading speed in peripheral vision. *Journal of Vision*, *13*(7):14, 1–14, https://doi.org/10.1167/13.7.14. [PubMed] [Article]

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., Zhao, Y., Lesmes, L., Bex, P., Yu, D. Y., & Lu, Z. L. (2018). Bayesian adaptive assessment of the reading function for vision: The qReading method. *Journal of Vision*, *18*(9):6, 1–15, https://doi.org/10.1167/18.9.6. [PubMed] [Article].

Kim, W., Pitt, M. A., Lu, Z. L., Steyvers, M., & Myung, J. I. (2014). A hierarchical adaptive approach to optimal experimental design. *Neural Computation*, *26*(11), 2465–2492, https://doi.org/10.1162/NECO_a_00654.

King-Smith, P. E., Grigsby, S. S., Vingrys, A. J., Benes, S. C., & Supowit, A. (1994). Efficient and unbiased modifications of the quest threshold method—Theory, simulations, experimental evaluation and practical implementation. *Vision Research*, *34*(7), 885–912, https://doi.org/10.1016/0042-6989(94)90039-6.

King-Smith, P. E., & Rose, D. (1997). Principles of an adaptive method for measuring the slope of the psychometric function. *Vision Research*, *37*(12), 1595–1604, https://doi.org/10.1016/S0042-6989(96)00310-0.

Klein, R., Wang, Q., Klein, B. E. K., Moss, S. E., & Meuer, S. M. (1995). The relationship of age-related maculopathy, cataract, and glaucoma to visual-acuity. *Investigative Ophthalmology & Visual Science*, *36*(1), 182–191.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3? *Perception*, *36*(14), 1.

Kontsevich, L. L., & Tyler, C. W. (1999). Bayesian adaptive estimation of psychometric slope and threshold. *Vision Research*, *39*(16), 2729–2737, https://doi.org/10.1016/S0042-6989(98)00285-5.

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

Legge, G. E. (2007). *Psychophysics of reading in normal and low vision*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

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., Cheung, S. H., Yu, D., Chung, S. T., Lee, H. W., & Owens, D. P. (2007). The case for the visual span as a sensory bottleneck in reading. *Journal of Vision*, *7*(2):9, 1–15, https://doi.org/10.1167/7.2.9. [PubMed] [Article]

Legge, G. E., Ross, J. A., Luebker, A., & Lamay, J. M. (1989). Psychophysics of reading. 8. The Minnesota Low-Vision Reading Test. *Optometry and Vision Science*, *66*(12), 843–853, https://doi.org/10.1097/00006324-198912000-00008.

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, 1–21, https://doi.org/10.1167/10.3.17. [PubMed] [Article]

Lu, Z. L., & Dosher, B. (2013). Adaptive psychophysical procedures. In *Visual psychophysics: from laboratory to theory* (pp. 351–384). Boston, MA: MIT Press.

MacKeben, M., Nair, U. K. W., Walker, L. L., & Fletcher, D. C. (2015). Random word recognition

chart helps scotoma assessment in low vision. *Optometry and Vision Science*, *92*(4), 421–428, https://doi.org/10.1097/Opx.0000000000000548.

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.

Owsley, C., McGwin, G., Lee, P. P., Wasserman, N., & Searcey, K. (2009). Characteristics of low-vision rehabilitation services in the United States. *Archives of Ophthalmology*, *127*(5), 681–689, https://doi.org/10.1001/archophthalmol.2009.55.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442, https://doi.org/10.1163/156856897x00366.

Pelli, D. G., & Tillman, K. A. (2007). Parts, wholes, and context in reading: A triple dissociation. *PloS One*, *2*(8), e680.

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]. *Klin Monbl Augenheilkd*, *213*(3), 174–181, https://doi.org/10.1055/s-2008-1034969.

Remus, J. J., & Collins, L. M. (2007). A comparison of adaptive psychometric procedures based on the theory of optimal experiments and Bayesian techniques: Implications for cochlear implant testing. *Perception & Psychophysics*, *69*(3), 311–323, https://doi.org/10.3758/Bf03193752.

Remus, J. J., & Collins, L. M. (2008). Comparison of adaptive psychometric procedures motivated by the theory of optimal experiments: Simulated and experimental results. *Journal of the Acoustical Society of America*, *123*(1), 315–326, https://doi.org/10.1121/1.2816567.

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

Rubin, G. S. (2013). Measuring reading performance. *Vision Research*, *90*, 43–51, https://doi.org/10.1016/j.visres.2013.02.015.

Rubin, G. S., & Turano, K. (1992). Reading without saccadic eye-movements. *Vision Research*, *32*(5), 895–902, https://doi.org/10.1016/0042-6989(92)90032-E.

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.

Runge, P. E. (2000). Eduard Jaeger's test-types (Schrift-Scalen) and the historical development of vision tests. *Transactions of the American Ophthalmological Society*, *98*, 375–438.

Seiple, W., Szlyk, J. P., Mcmahon, T., Pulido, J., & Fishman, G. A. (2005). Eye-movement training for reading in patients with age-related macular degeneration. *Investigative Opthalmology & Visual Science*, *46*(8), 2886–2896, https://doi.org/10.1167/iovs.04-1296.

Sloan, L. L., & Brown, D. J. (1963). Reading cards for selection of optical aids for the partially sighted. *American Journal of Ophthalmology*, *55*(6), 1187–1199, https://doi.org/10.1016/0002-9394(63)90188-0.

Snoeren, P. R., & Puts, M. J. H. (1997). Multiple parameter estimation in an adaptive psychometric method: MUEST, an extension of the QUEST method. *Journal of Mathematical Psychology*, *41*(4), 431–439, https://doi.org/10.1006/jmps.1997.1188.

Sunness, J. S., Rubin, G. S., Applegate, C. A., Bressler, N. M., Marsh, M. J., Hawkins, B. S., & Haselwood, D. (1997). Visual function abnormalities and prognosis in eyes with age-related geographic atrophy of the macula and good visual acuity. *Ophthalmology*, *104*(10), 1677–1691, https://doi.org/10.1016/S0161-6420(97)30079-7.

Watson, A. B. (2017). QUEST plus: 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]

Watson, A. B., & Pelli, D. G. (1983). Quest—A Bayesian adaptive psychometric method. *Perception & Psychophysics*, *33*(2), 113–120, https://doi.org/10.3758/Bf03202828.

Whittaker, S. G., & Lovie-Kitchin, J. (1993). Visual requirements for reading. *Optometry and Vision Science*, *70*(1), 54–65, https://doi.org/10.1097/00006324-199301000-00010.

World Health Organization. (2014). Visual impairment and blindness. *Fact Sheet No. 282*. Geneva: WHO. Retrieved from http://www.who.int/mediacentre/factsheets/fs282/en/.

Yu, D., 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]

Yu, D., Legge, G. E., Wagoner, G., & Chung, S. T. (2014). Sensory factors limiting horizontal and

vertical visual span for letter recognition. *Journal of Vision*, *14*(6):3, 1–17, https://doi.org/10.1167/14.6.3. [PubMed] [Article]

Yu, D., Park, H., Gerold, D., & Legge, G. E. (2010). Comparing reading speed for horizontal and vertical English text. *Journal of Vision*, *10*(2):21, 1–17, https://doi.org/10.1167/10.2.21. [PubMed] [Article]