Methods

Psychophysical Validation of a Novel Active Learning Approach for Measuring the Visual Acuity Behavioral Function

Yukai Zhao[1], Luis Andres Lesmes[2], Michael Dorr[2,3], Peter J. Bex[4], and Zhong-Lin Lu[1,5–7]

[1] Center for Neural Science, New York University, New York, NY, USA
[2] Adaptive Sensory Technology, San Diego, CA, USA
[3] Technical University of Munich, Munich, Germany
[4] Department of Psychology, Northeastern University, Boston, MA, USA
[5] Division of Arts and Sciences, NYU Shanghai, Shanghai, China
[6] Department of Psychology, New York University, New York, NY, USA
[7] NYU-ECNU Institute of Brain and Cognitive Neuroscience, Shanghai, China

Purpose: To evaluate the performance of the quantitative visual acuity (qVA) method in measuring the visual acuity (VA) behavioral function.

Methods: We evaluated qVA performance in terms of the accuracy, precision, and efficiency of the estimated VA threshold and range in Monte Carlo simulations and a psychophysical experiment. We also compared the estimated VA threshold from the qVA method with that from the Electronic Early Treatment Diabetic Retinopathy Study (E-ETDRS) and Freiburg Visual Acuity Text (FrACT) methods. Four repeated measures with all three methods were conducted in four Bangerter foil conditions in 14 eyes.

Results: In both simulations and psychophysical experiment, the qVA method quantified the full acuity behavioral function with two psychometric parameters (VA threshold and VA range) with virtually no bias and with high precision and efficiency. There was a significant correlation between qVA estimates of VA threshold and range in the psychophysical experiment. In addition, qVA threshold estimates were highly correlated with those from the E-ETDRS and FrACT methods.

Conclusions: The qVA method can provide an accurate, precise, and efficient assessment of the full acuity behavioral function with both VA threshold and range.

Translational Relevance: The qVA method can accurately, precisely, and efficiently assess the full VA behavioral function. Further research will evaluate the potential value of these rich measures for both clinical research and patient care.

Introduction

Visual acuity (VA) is a measure of spatial resolution that provides the most important metric in assessing functional vision.[1] As a primary clinical measure for characterizing optical and neural deficits and an important endpoint for treatment efficacy in a large number of diseases,[2] VA is also the metric for specific minimum vision standards in many professions.[3] For these reasons, the quality of visual acuity assessment is extremely important. Inaccurate and/or imprecise

VA assessment could result in unfair classification for paralympic athletes,[4] lost job opportunities,[5,6] or missed diagnoses of real disease and its progression, which may lead to loss of disability benefits[7] or delayed treatment.[8]

VA is typically expressed as a single score in logMAR, obtained from testing vision with printed VA charts[9–16] and computerized tests.[17–25] However, a single VA score alone does not describe the full visual acuity behavior of the observer; rather, an acuity psychometric function (i.e., performance in optotype recognition as a function of optotype size) is required
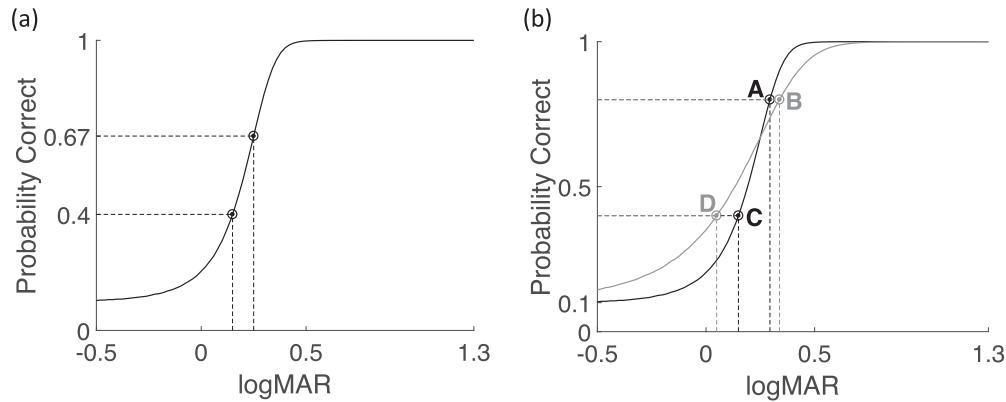
Figure 1. Visual acuity psychometric function. (a) A single VA psychometric function; the VA thresholds at two different performance levels are different. (b) Two VA psychometric functions with different slopes; changes in the VA thresholds at two different performance levels exhibit opposite signs. The values 0.02 logMAR and 0.10 logMAR correspond to a VA change of one letter and one line on an ETDRS chart, respectively.

(Fig. 1a). Typically, the function has a sigmoidal shape, is monotonically increasing with optotype size, and is well characterized by a mathematical formula with two parameters: threshold and slope (or range), with the VA threshold defined as the optotype size corresponding to a defined performance level (e.g., 67% correct) and slope (or range) quantifying the steepness of the psychometric function—that is, how fast acuity behavior changes with increasing or decreasing optotype sizes. Both VA threshold and slope may vary across individuals and disease stages.[26–29] For an observer described by a single acuity psychometric function, VA thresholds at different performance levels are different (Fig. 1a); we cannot directly compare estimated VA scores of an observer obtained from different instruments unless they measure VA thresholds at the same performance level. Moreover, when we compare different observers or a single observer at different disease stages described by acuity psychometric functions with different slopes (Fig. 1b), the magnitude and sign of change in VA threshold depend on the performance level. It may increase at one performance level (B–A in Fig. 1b) but decrease at another performance level (D–C in Fig. 1b). Specifications of the VA threshold with its corresponding target performance level and the slope (or range) of the acuity psychometric function are both necessary to interpret VA and VA changes.

Unfortunately, printed VA charts[9–16] and computerized tests,[17–25] except for the Freiburg Visual Acuity Text (FrACT) method,[30] use operational procedures to generate VA scores without specifying the VA behavioral function, the target performance level, and the slope of the acuity psychometric function. For the same observer, VA scores obtained from different charts and/or the same chart with different heuristic termination rules can be different[31] because they target

different performance levels on the acuity psychometric function (Fig. 1a); yet, an unbiased estimate of VA change can only be obtained when the two VA estimates are obtained at the same performance level. For different observers measured with the same chart and heuristic termination rule, or a single observer tested at different disease stages, VA scores may correspond to different performance levels if the VA behavioral functions have different slopes (Fig. 1b). To compare VA scores obtained from these tests, the slope (or range) of the VA behavioral function is necessary for correcting differences caused by different performance levels. However, none of the existing VA instruments[17–25,30] measures the slope of the VA behavioral function.

In addition, existing tests face a challenge of achieving high accuracy, precision, and efficiency. The most popular VA chart, the Snellen chart,[16] is neither accurate nor precise.[14,32–35] Following Bailey and Lovie,[10] logMAR charts[14,17–19,33,36–38] and computerized tests[17–25] were developed to improve the precision of VA testing. However, the improvement has been limited because of the coarse sampling density of optotype size required to limit the library of test items to reduce testing time.[14,33] This is true even for the Early Treatment Diabetic Retinopathy Study (ETDRS) chart[12] (Fig. 2a) and its computerized version, the E-ETDRS method,[17] with 0.02 logMAR and 0.10 logMAR corresponding to a VA change of one letter and one line on an ETDRS chart, respectively. More recently, new logMAR charts and computerized tests have been developed to either reduce testing time at the cost of precision or improve precision at the cost of testing time.[14,18,30,33,39,40] However, it is still a challenge to assess VA with both high precision and high efficiency.
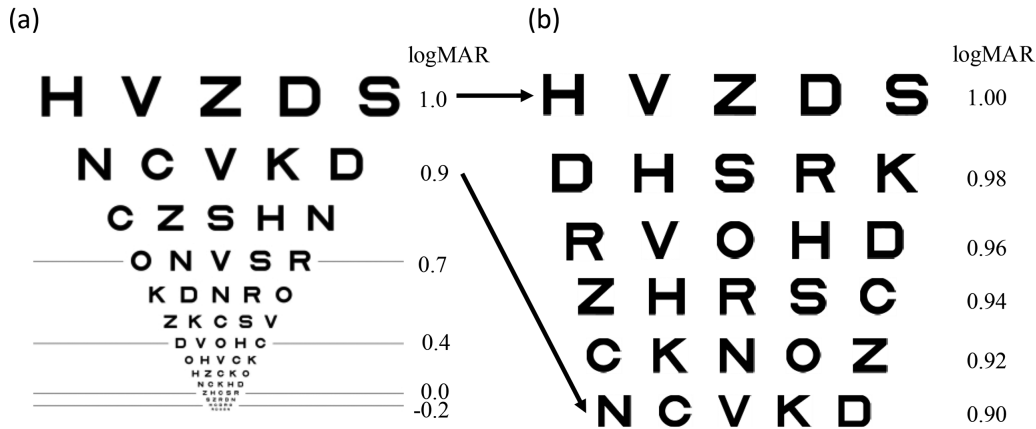
Figure 2. Optotype size sampling densities of (a) ETDRS (0.10 logMAR between rows) and (b) qVA (0.02 logMAR between rows).

Recently, a novel quantitative visual acuity (qVA) test (Patent No. US 10758120B2; Lesmes LA. *IOVS*. 2018;59:ARVO E-Abstract 1073)[41] was developed to characterize the full VA behavioral function using a combination of Bayesian active learning and high-density sampling of optotype size (Fig. 2b; Supplementary Material, Part A). The qVA method models performance in identifying multiple optotypes in each trial with a VA behavioral function with two parameters: VA threshold at a fixed d' performance level and VA range, corresponding to the steepness of the VA behavioral function (Supplementary Fig. S1). To sample the continuum of optotype size with fine-grained resolution and at the same time achieve high efficiency, the qVA implements an active learning approach to optimize the test stimuli in each trial.[42–61] Specifically, the testing algorithm begins with a joint prior probability distribution of VA threshold and range of the acuity psychometric function, selects the most informative test stimuli based on all available information before each trial, and updates the posterior distribution with Bayes' rule. The result is a joint posterior distribution (Supplementary Fig. S2) that can be used to estimate VA threshold and VA range, as well as their uncertainties.

In this proof-of-concept study, we used both computer simulations and a psychophysical experiment to evaluate qVA performance in terms of the accuracy, precision, and efficiency for estimating VA threshold and VA range. In addition, we compare estimated VA thresholds obtained from qVA with those from E-ETDRS and FrACT methods. The qVA method, like the E-ETDRS and the FrACT methods, is an optotype test. In this study, we applied the qVA method with three optotypes per row in both simulations and the psychophysical experiment and used the same Sloan optotype set used by the other two methods. The results

suggest that qVA can estimate the full acuity behavioral function—both threshold and range—with accurate, precise, and efficient assessment.

Simulations

Methods

Acuity Tests

The details of the qVA (Patent No. US 10758120B2; Lesmes LA. *IOVS*. 2018;59:ARVO E-Abstract 1073),[41] E-ETDRS,[17] and FrACT[26,30] methods have been published in the cited references and are briefly summarized in the Supplementary Material, Parts A, B, and C and Figures S1, S2, S3, S4, and S5. Whereas the qVA and FrACT methods measure the VA threshold at 67% and 55% correct, respectively, the E-ETDRS method uses an operational procedure that can lead to VA threshold estimates at different performance levels for observers with different slopes of the acuity behavioral function.

Simulated Observers

We conducted Monte Carlo simulations using the qVA,[41] E-ETDRS,[17] and FrACT[26,30] methods in a 10-alternative forced-choice (10AFC) letter identification task for two observers with the same VA threshold—0.25 logMAR at $d' = 2$ (i.e., 67% correct in 10AFC optotype identification)—but different ranges (0.3 and 0.6 logMAR). Specifically, the performance of each simulated observer in identifying multiple optotypes (probabilities of correctly identifying k [$= 0, …, N$] out of N optotypes) in each trial was modeled by composite multiple-optotype psychometric functions based on the single-optotype d' acuity psychometric function (Supplementary Material, Part A).

Simulation Procedures

One thousand qVA runs with 45 three-optotype rows (135 optotypes), 1000 E-ETDRS runs, and 1000 FrACT runs of 45 optotypes were simulated for each observer.

For each method, the N optotypes presented in each trial were determined by the same test procedure for real observers (Supplementary Material, Parts A–C). The only difference between the simulation and the psychophysical testing was that the response of each simulated observer was determined by an acuity psychometric function, which specifies the probability of correctly identifying k (= 0, …, N) out of N optotypes of a given size, along with a random number that was used to select the actual number of correct responses in each trial.

Reanalysis of E-ETDRS and FrACT with Bayesian Procedure in the qVA Algorithm

We used the simulated E-ETDRS and FrACT data as the input and used the scoring algorithm in the qVA to estimate both VA threshold and VA range. In this reanalysis, the sequence of optotype sizes was selected by either the E-ETDRS or the FrACT method, and the responses were made by the simulated observers in the tests, but the data were scored with the Bayesian procedure in the qVA algorithm.

Evaluations

The estimated parameters from VA tests reflect the contributions of both systematic and random measurement errors.[62] The systematic measurement error or bias is defined as the difference between the mean of an estimated parameter across (infinitely) repeated measurements and the truth (Fig. 3). The smaller the bias is, the more accurate the estimate. In empirical studies, the truth is often unknown; the agreement between one method and a known standard method, evaluated by the 95% limits of agreement between the two methods in the Bland–Altman analysis,[63,64] is typically used to provide a proxy measure of accuracy. When the two methods cannot be directly compared, correlation is computed instead. The random measurement error refers to the variability of repeated measures of an estimated parameter. It is typically quantified by the standard deviation (SD) of the estimated parameter from repeated measures with a single method or the 95% repeatability coefficient[64] that is defined as $1.96 \times \sqrt{2}$SD. For continuous measurements, precision is often quantified with 1/SD; however, coarse measurements with large quantization steps may lead to artificially small SDs that do not necessarily imply good precision.[65] Fractional Rank Precision (FRP) was recently introduced as a metric that takes into account
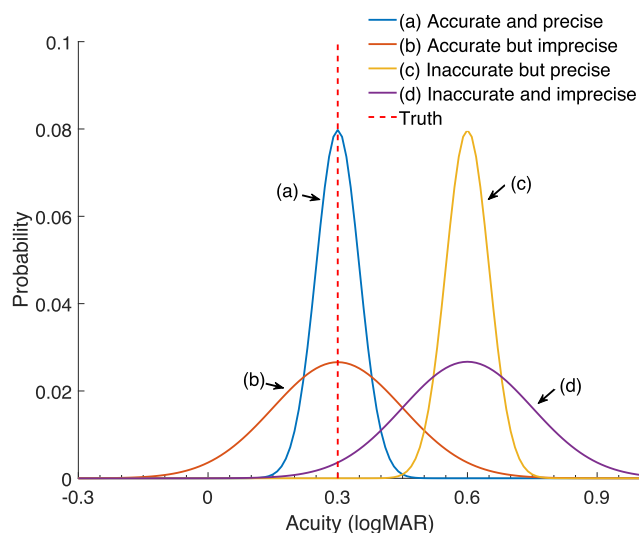

Figure 3. Illustration of accuracy and precision. The true acuity is 0.3 logMAR. The means of the accurate (a, b) and inaccurate (c, d) measurements are 0.3 and 0.6 logMAR, respectively. The SDs of the precise (a, c) and imprecise (b, d) measurements are 0.05 and 0.15 logMAR, respectively.

the effects of measurement grain.[66] In addition, the testing time and/or number of optotypes required to reach a desired accuracy and/or precision level are used to quantify the efficiency of a method.

The qVA method generated a joint posterior distribution of the two parameters of the VA behavioral function (VA threshold and VA range) after each trial (Supplementary Material, Part D). The accuracy of each estimated parameter was then quantified by its bias. The precision was quantified by both cross-run and within-run variability. The cross-run variability of each estimated parameter was quantified by its SD across the 1000 simulated runs (tests). The within-run variability of each estimated parameter was quantified by the half-width of the 68.2% credible interval (68.2% HWCI) of its marginal posterior distribution.[67,68] The 68.2% HWCI defines the shortest interval that contains the true value of the estimated quantity with 68.2% probability (for details, see Supplementary Material, Part D). The 68.2% HWCI is equal to the cross-run variability (SD) under two conditions: (1) observer behavior does not change across runs, and (2) the qVA procedure has converged with sufficient testing.

The bias and precision of the estimated VA from the FrACT method were defined similarly, except that the true VAs of the simulated observers were set at 0.210 and 0.170 logMAR, respectively, corresponding to the 55% correct performance level. For the E-ETDRS method, only the precision of the estimated VA scores was evaluated. We could not compute the
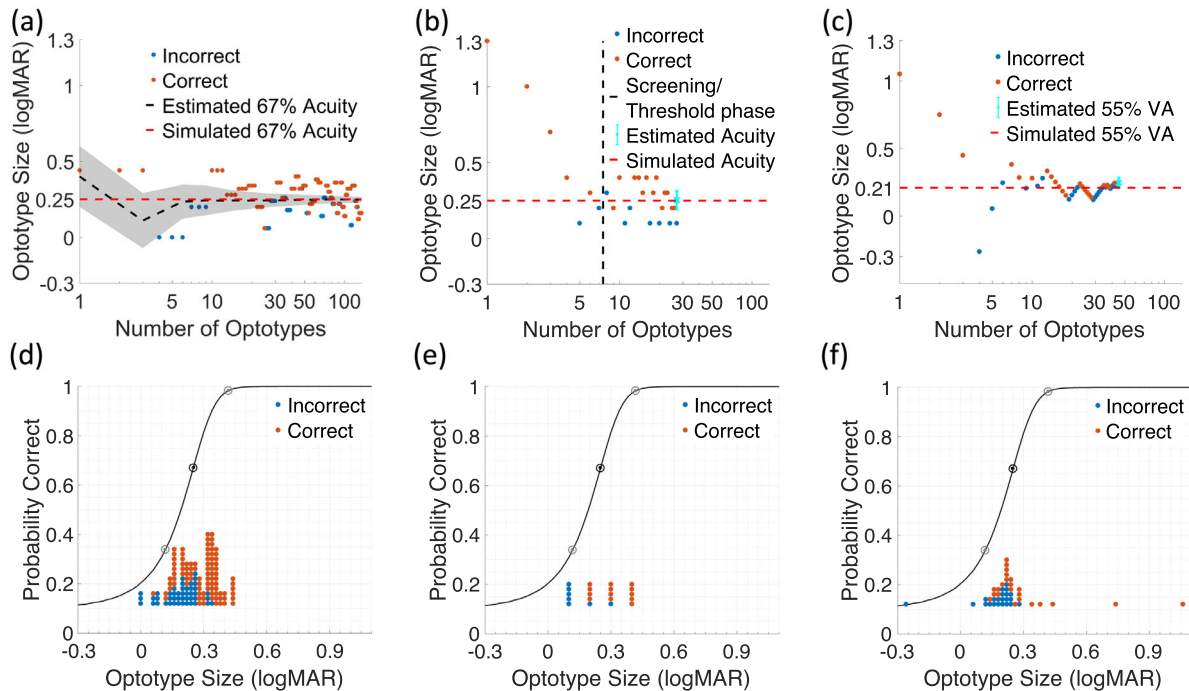
Figure 4. (a, b, c) Example test histories of the qVA (a), E-ETDRS (b), and FrACT (c) methods in the simulation study (Observer 1). The true acuity of the simulated observer is represented by the horizontal *dashed red line*. The color of each dot indicates correct/incorrect responses. In a, the estimated row-by-row VA threshold and its SD are represented by the *dashed black lines* and *shaded grey areas*, respectively. In b, the vertical *dashed black line* separates the screening and the threshold phases. In b and c, the estimated VA threshold and its SD are represented by the *light blue cross* and its error bar. (d, e, f) The stimulus placement from a, b, and c relative to the acuity psychometric function; for the E-ETDRS method, only the stimuli in the threshold phase are shown in e.

true VA scores and therefore the bias because the method does not specify the target performance level.

We computed the sensitivity at 95% specificity to detect a true change of 0.15 logMAR. Given a change criterion and results from two repeated measures with a single method, specificity is defined as the probability of correctly identifying an individual who has not undergone a change, and sensitivity is defined as the probability of correctly identifying an individual who has undergone a change.[69,70] Using the 95% repeatability coefficient[64] (1.96 × $\sqrt{2}$SD) as the change criterion,[69,70] the specificity is 95% by definition. Sensitivity at 95% specificity is equal to P{X > 1.96 × $(1 - \frac{d}{1.96 \times \sqrt{2}SD})$)} where X is a random variable, normally distributed with mean = 0 and SD = 1; d = 0.15 logMAR is the true VA change.

We chose the change criterion corresponding to specificity at 95% to compare sensitivity across different methods on an equal footing.[69,70] Although the logMAR values associated with the criterion level changed with the SD of the estimated VA, the criterion itself was the same if it is normalized by the SD, and the specificity was the same (95%) across all the methods

and conditions. We chose 0.15 logMAR as the magnitude of true change, because a greater-than-15-letter VA improvement (0.3 logMAR) is considered by the US Food and Drug Administration as an acceptable endpoint of a clinical trial, although a greater-than-10-letter VA improvement (0.2 logMAR) has been used when the benefits can outweigh the safety risks of the proposed method or product.[71]

Results

Figure 4 shows the test histories of the qVA (Figs. 4a, 4d), E-ETDRS (Figs. 4b, 4e), and FrACT (Figs. 4c, 4f) methods in one run of simulated Observer 1.

Estimated VA and Range from the qVA Method

The qVA method provided accurate (Figs. 5, 6) and precise (Figs. 7, 8) estimates of both VA threshold and range (Table 1). After 135 optotypes, the bias of the VA threshold and range estimates were –0.001 and 0.017 logMAR, respectively, for Observer 1, and 0.000 and –0.018 logMAR, respectively, for Observer 2. After
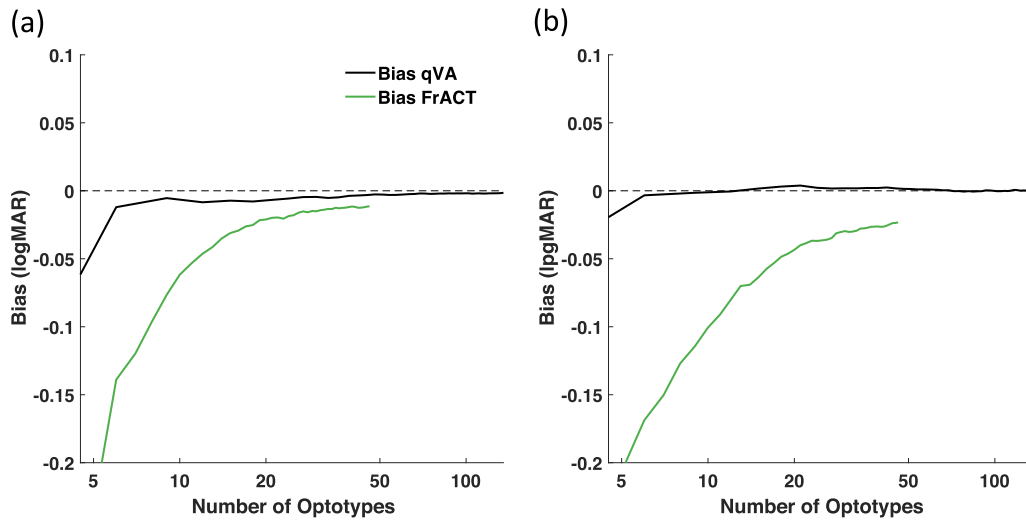
Figure 5. Bias of the estimated VA from the qVA (*black lines*) and FrACT (*green lines*) methods as functions of number of optotypes for simulated Observer 1 (a) and Observer 2 (b). The horizontal *dashed lines* indicate zero bias.

Figure 6. Bias of the estimated range from the qVA (*black lines*) method as functions of number of optotypes for simulated Observer 1 (a) and Observer 2 (b). The horizontal *dashed lines* indicate zero bias.
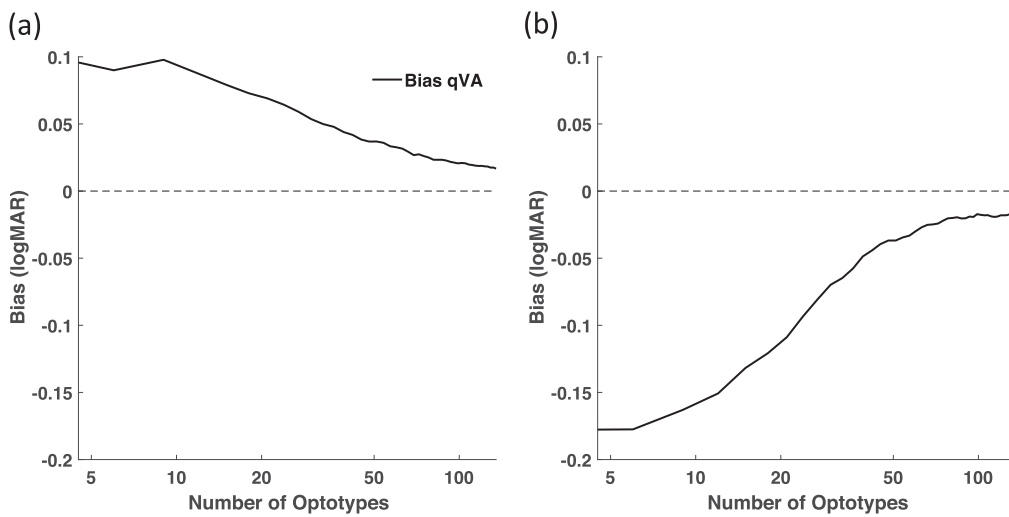
135 optotypes, the SDs of the VA threshold and range estimates were 0.020 and 0.060 logMAR, respectively, for Observer 1 and were 0.037 and 0.117 logMAR, respectively, for Observer 2. The 68.2% HWCI of the VA threshold and range estimates were 0.020 and 0.066 logMAR, respectively, for Observer 1 and were 0.036 and 0.119 logMAR, respectively, for Observer 2. The results suggest that the qVA provides both accurate and precise estimates of the VA behavioral function: both VA threshold and range.

Comparison of Estimated VA from the Three Methods

Because the qVA provides both VA threshold and range estimates, we compared the VA estimates

obtained from the qVA with those from the E-ETDRS and FrACT methods.

The VA thresholds obtained from the qVA exhibited higher accuracy (bias = –0.001 and 0.000 logMAR for Observers 1 and 2, respectively) than those from FrACT (bias = –0.011 and –0.023 logMAR for Observers 1 and 2, respectively). We cannot compute the bias of the estimated VA from the E-ETDRS method because the cumulative letter score does not specify the target performance level.

The estimated VA from the qVA method also exhibited higher precision (SD = 0.020 and 0.037 logMAR for Observers 1 and 2, respectively) than those from the FrACT (SD = 0.038 and 0.063 logMAR for Observers
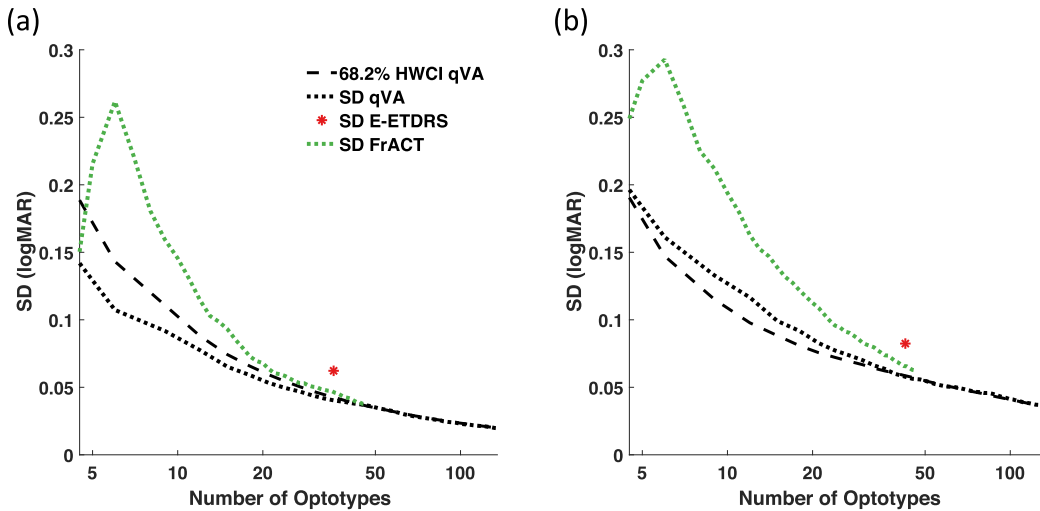
Figure 7. SDs of the estimated VA from the qVA (68.2% HWCI, *black dashed lines*; SDs, *black dotted lines*), E-ETDRS (*red asterisks*), and FrACT (*green dotted lines*) methods as functions of number of optotypes for simulated Observer 1 (a) and Observer 2 (b).
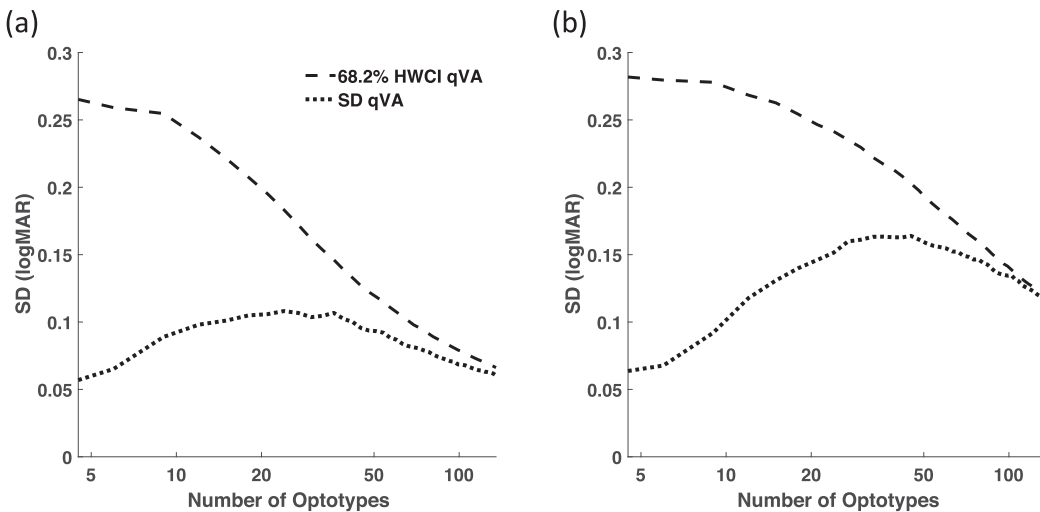

Figure 8. SDs of the estimated range from the qVA method as functions of number of optotypes for simulated Observer 1 (a) and Observer 2 (b) (68.2% HWCI, *black dashed lines*; SDs, *black dotted lines*).

1 and 2, respectively) and E-ETDRS (SD = 0.062 and 0.082 logMAR for Observers 1 and 2, respectively) methods (Fig. 9). At 95% specificity, the sensitivity values for detecting an estimated change of 0.15 logMAR were 100.0%, 40.2%, and 79.7% for Observer 1 and were 81.8%, 25.3%, and 39.1% for Observer 2 for the qVA, FrACT, and E-ETDRS methods, respectively (Table 2).

The qVA method was also more efficient than the other two methods in estimating VA threshold. It required only nine and six optotypes for Observers 1 and 2, respectively, to reach the same bias (–0.011 and –0.023 logMAR) and 42 and 39 optotypes to reach the same precision (SD = 0.038 and 0.063 logMAR for Observers 1 and 2, respectively) of the FrACT method,

which was based on 45 optotypes. It required 18 and 24 optotypes to reach the same precision of the E-ETDRS method (SD = 0.062 and 0.082 logMAR), which was based on 36 and 43 optotypes (on average) for Observers 1 and 2, respectively.

Reanalysis of the E-ETDRS and FrACT Methods

The reanalysis did improve the accuracy and precision of the estimated VA in some cases (Tables 1, 2). The SDs of the reanalyzed VA from the E-ETDRS simulations were reduced by 25.8% and 7.3% for simulated Observers 1 and 2, respectively, and the bias of the reanalyzed VA from the FrACT simulations was reduced by 54% for simulated Observer 2. On the other hand, the estimated slope (quantified

Table 1. Bias, SDs, and 68.2% HWCI of the Estimates from the Three Methods in the Simulations

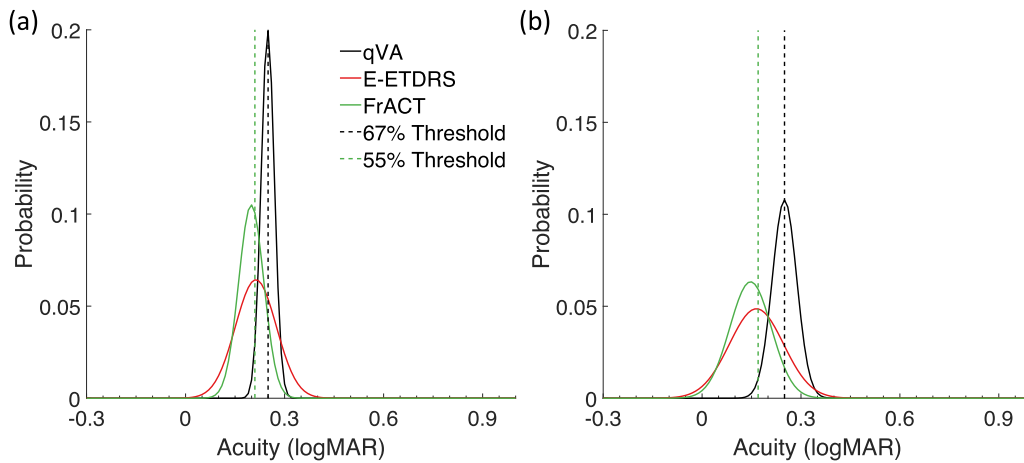|			Observer			
			1	2	1	2
Statistic	Method	Number of Optotypes	Acuity		Range	
Bias	qVA	36, 42	−0.005	0.002	0.048	−0.044
		45	−0.003	0.002	0.038	−0.040
		135	−0.001	0.000	0.017	−0.018
	FrACT	45	−0.011	−0.023	N/A	N/A
	Reanalyzed E-ETDRS	36, 42	0.008	−0.006	0.045	−0.078
	Reanalyzed FrACT	45	0.019	−0.012	0.071	−0.133
SD	qVA	36, 42	0.040	0.058	0.107	0.163
		45	0.037	0.056	0.095	0.164
		135	0.020	0.037	0.060	0.117
	E-ETDRS	36, 42	0.062	0.082	N/A	N/A
	FrACT	45	0.038	0.063	N/A	N/A
	Re-analyzed E-ETDRS	36, 42	0.046	0.076	0.103	0.178
	Re-analyzed FrACT	45	0.037	0.066	0.091	0.139
68.2% HWCI	qVA	36, 42	0.042	0.059	0.146	0.208
		45	0.037	0.057	0.127	0.203
		135	0.020	0.036	0.066	0.119
	Re-analyzed E-ETDRS	36,42	0.052	0.066	0.153	0.206
	Re-analyzed FrACT	45	0.049	0.062	0.204	0.237

Figure 9. Distributions of the estimated VA threshold from the qVA (*black lines*), E-ETDRS (*red lines*), and FrACT (*green lines*) methods for simulated Observer 1 (a) and Observer 2 (b).

by VA range) did not converge in the reanalysis, as indicated by the larger 68.2% HWCIs relative to the SDs. In general, the estimated VA from the reanalysis of the simulated E-ETDRS and FrACT data was less accurate and precise than that from the qVA when the number of optotypes was matched, especially for simulated Observer 2, who had a shallower acuity psychometric function. The estimated VA range from

the simulated FrACT data exhibited the largest bias (86% and 233% greater than those from the qVA) because the FrACT test concentrated on optotype sizes near the 55% correct VA threshold. In summary, the qVA method provided the most accurate and precise estimates of both VA threshold and VA range parameters, even when the number of optotypes was matched.

Table 2. Sensitivity at 95% Specificity to Detect an Estimated Change of 0.15 logMAR for the Three Methods in the Simulations

Observer	Method	Number of Optotypes	SD	Sensitivity (%) at 95% Specificity
1	qVA	36, 42	0.040	75.5
		45	0.037	81.8
		135	0.020	100.0
	E-ETDRS	36, 42	0.062	40.2
	FrACT	45	0.038	79.7
	Reanalyzed E-ETDRS	36, 42	0.046	63.5
	Reanalyzed FrACT	45	0.037	81.8
2	qVA	36, 42	0.058	44.8
		45	0.056	47.4
		135	0.037	81.8
	E-ETDRS	36, 42	0.082	25.3
	FrACT	45	0.063	39.1
	Reanalyzed E-ETDRS	36, 42	0.076	28.6
	Reanalyzed FrACT	45	0.066	36.2

Psychophysical Evaluation

We conducted a psychophysical experiment to evaluate the precision of VA threshold and range estimates obtained from qVA and compared its estimated VA threshold with those from E-ETDRS and FrACT. Pairwise correlations between VA estimates from the three methods were used to validate the qVA method. Although all subjects had normal or corrected-to-normal vision, we degraded their vision with three different levels of Bangerter foils[72] to expand the dynamic range of their measurable visual acuities.

Methods

Apparatus

The experiment was conducted on a PC computer with Psychtoolbox 3.0.11 extensions[73] in MATLAB R2013a (MathWorks, Natick, MA). The computer was used to drive a 24-inch Dell P2415Q liquid-crystal display monitor (Dell Technologies, Round Rock, TX) with a resolution of 3840 × 2160 pixels and a background luminance of 97 cd/m^2. The display was viewed monocularly with natural pupil at a viewing distance of 4 meters while the other eye was covered by an opaque eye patch. A chin rest was used to stabilize the observer's head. Acuity was degraded by three levels of Bangerter occlusion foils (Ryser Ophtalmologie, St. Gallen, Switzerland) with nominal acuities of 20/25, 20/30, and 20/100. The room was dimly lit throughout the experiment.
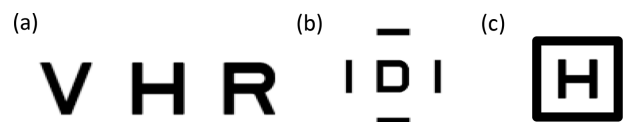

Figure 10. Illustration of test stimuli used in the qVA (a), E-ETDRS (b), and FrACT (c) methods.

Stimuli

The optotype stimuli consisted of 10 black Sloan letters (C, D, H, K, N, O, R, S, V, and Z). In the qVA method, Sloan letters[74] in 91 optotype sizes, equally spaced between –0.5 and 1.3 logMAR with step-size resolution of 0.02 logMAR, were used. Three letters of the same size and with a center-to-center distance of 1.75 letter width were presented in each qVA trial (one row) (Fig. 10a). In the E-ETDRS method, Sloan letters[74] of 20 optotype sizes, equally spaced between –0.3 and 1.6 logMAR with a step size of 0.1 logMAR, were used. To mimic crowding effects in the ETDRS chart in which each row consists of five optotypes, flanker bars with the same stroke width as the letter and a flanker-to-center distance of 1.75 letter width were used (Fig. 10b). In the FrACT method, Sloan letters with optotype sizes sampled between –0.76 and 1.11 logMAR were used along with an anti-aliasing method.[75] A two-letter-wide box was used to mimic crowding effects (Fig. 10c).

Observers

Six naïve observers and the first author participated in the experiment (25–48 years old; mean, 35 years; all with at least a bachelor's degree). All observers had normal or corrected-to-normal visual acuity (equal to
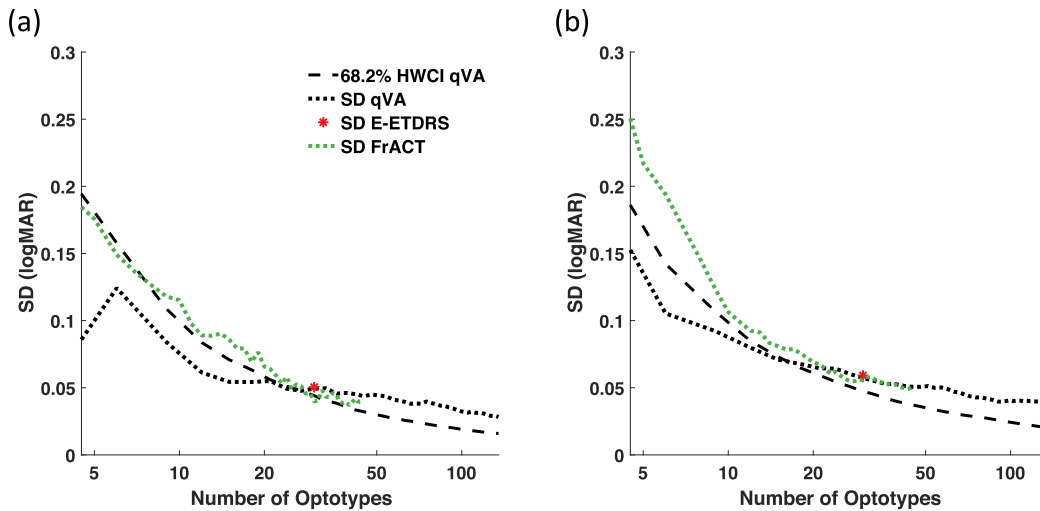
Figure 11. SDs of the estimated VA from the qVA (68.2% HWCI, *black dashed lines*; SDs, *black dotted lines*), E-ETDRS (*red asterisks*), and FrACT (*green dotted lines*) methods as functions of number of optotypes in the no-foil condition (a) and three foil conditions (b), with between-block variability correction, in the psychophysical experiment.
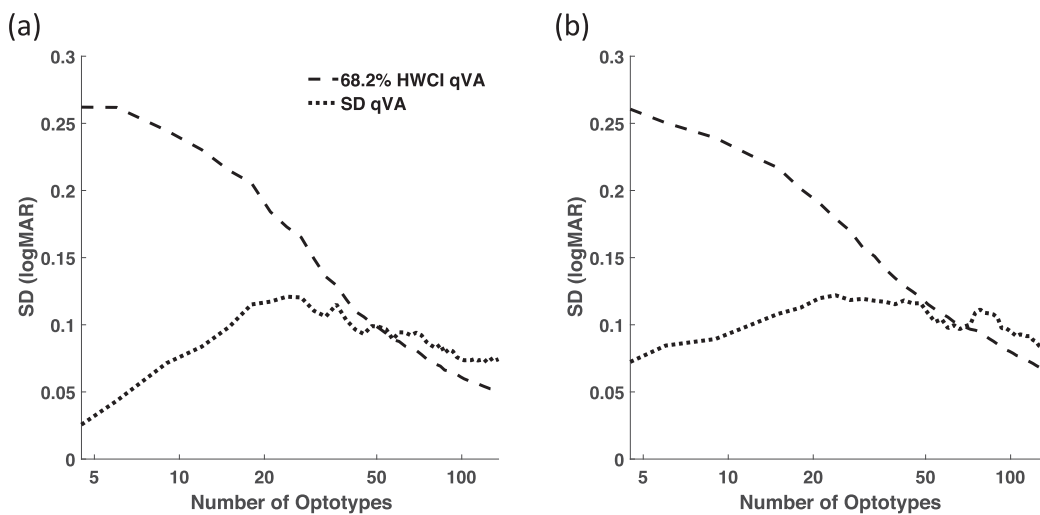

Figure 12. SDs of the estimated range from the qVA (68.2% HWCI, *black dashed lines*; SDs, *black dotted lines*) method as functions of number of optotypes in the no-foil condition (a) and three foil conditions (b) in the psychophysical experiment.

or better than 0.0 logMAR). The study was approved by the Institutional Review Board of Human Subject Research at the Ohio State University. Written consent was obtained from each observer in the beginning of the experiment. Observers (except the first author) were paid $10/hour. The research followed the tenets of the Declaration of Helsinki.

Design

Monocular acuity behavior, with the non-tested eye covered by an eye patch, was measured for both eyes of each subject in four sessions, one for each of the three foil and no-foil conditions. Each session consisted of eight blocks, four for the left eye and four for the right eye. In each block, the subject was tested in one eye once with each of the three methods. The test order of the three methods and eyes was randomly selected for each block. This resulted in four repeated tests for each combination of subject, eye, method, and foil condition.

Procedure

Both the qVA and E-ETDRS methods were programmed in MATLAB. In each qVA trial, three letters on a row were randomly sampled without replacement from the 10 Sloan letters, with optotype size determined by the qVA method. The observer was instructed to report the identity of the three letters

Table 3. SDs of VA Estimates Obtained from Three Psychophysical Methods

Method	Number of Optotypes	No Foil	F1	F2	F3	Three Foil Average
				SD		
qVA	30	0.049	0.056	0.060	0.057	0.057
	45	0.044	0.053	0.054	0.045	0.051
	78	0.039	0.048	0.045	0.034	0.043
	135	0.028	0.049	0.033	0.034	0.039
E-ETDRS	30	0.050	0.049	0.060	0.067	0.059
FrACT	45	0.039	0.062	0.057	0.026	0.051

verbally from left to right. Each qVA run consisted of 45 three-optotype rows each (Fig. 10a; Supplementary Material, Part A). The E-ETDRS procedure[17] was used to program the E-ETDRS method (Fig. 10b; Supplementary Material, Part B). To accelerate testing, subjects were allowed to report "I don't know" in the qVA and E-ETDRS methods if a letter could not be identified, rather than forcing a response from the optotype set, and the corresponding response was scored as incorrect. FrACT software (version 3.9.8; downloaded from https://michaelbach.de/fract/) was used to test the FrACT method.[30] Each FrACT run consisted of 45 one-optotype trials with the "no easy trial" option (Fig. 10c; Supplementary Material, Part C). The experimenter entered the verbal responses via a wireless keyboard for all three methods.

Analyses

The precision of the estimated VA threshold and VA range from the qVA methods was quantified by the SD, defined as the SD of the four repeated measures of the same eye in the same condition. The average SD of each estimated parameter was defined as the square root of the mean variance across eyes.

Because the no-foil and foil conditions simulate vision in healthy and degraded vision, the SDs of the VA estimates were calculated in the no-foil and foil conditions separately. Although the Bangerter occlusion foils are labeled with nominal degrees of acuity degradation (e.g., 20/30), the optical degradations caused by them are known to deviate from the labels.[72,76] Furthermore, the irregular transparencies on the surface of the foils may lead to different degrees of visual degradation for the same observer wearing the same foil, depending on the exact alignment used at the time of testing.[76] We developed a method (Supplementary Material, Part E) to correct for the systematic between-block variability introduced by foil variabilities.

Because it provides both VA threshold and range estimates, we compared VA threshold estimates from qVA with those obtained from E-ETDRS and FrACT. To validate qVA, pairwise correlations were computed among the VA thresholds obtained from the three methods. Based on the estimated VA threshold and slope from the qVA method, we simulated the psychophysical experiment 100 times to compute the mean and SD of the pairwise correlation coefficients to estimate their upper bounds. SDs of the estimated VA from the E-ETDRS and FrACT methods were computed. Because SDs of repeated measures could be confounded by sampling resolution (or minimum step size) of the test,[65] FRP[66,77] was also used to evaluate test–retest precision of the estimated VA from the three methods.

Results

Testing Time

The mean testing time for one run of each test was 186 ± 29 seconds for qVA (135 optotypes), 66 ± 16 for E-ETDRS (30 optotypes), and 112 ± 25 seconds for FrACT (45 optotypes). It took on average 1.4, 2.2, and 2.5 seconds to test one optotype in the three methods, respectively. The average testing times per optotype were significantly different across the three methods, $F(2, 549) = 262.8$, $P < 0.001$, with that of qVA being significantly shorter than that of FrACT ($t = -22.63$, $P_{tukey} < 0.001$) and E-ETDRS ($t = -15.43$, $P_{tukey} < 0.001$). By using three optotypes in a row, qVA sped up testing time per optotype.

Estimated VA and Range from the qVA Method

The estimated VA threshold and range from the qVA method are shown in Supplementary Figures S6 and S7. The mean estimated ranges were 0.254, 0.340, 0.315, and 0.320 logMAR in the no-foil and three foil conditions, respectively, $F(3, 18) = 9.111$, $P < 0.001$. In addition, we found a significant correlation between the estimated VA threshold and range ($r = 0.412$, $P < 0.001$).
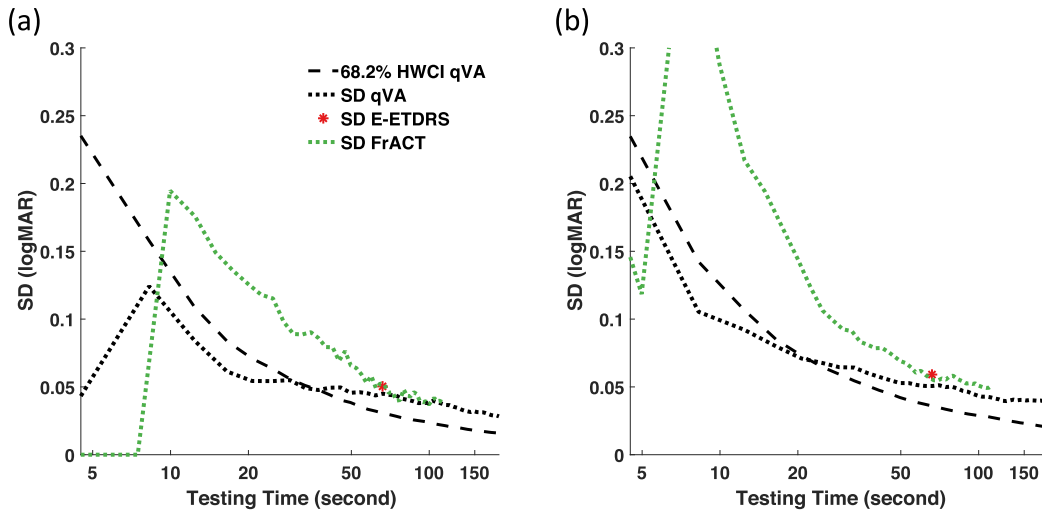
Figure 13. SDs of the estimated VA from the qVA (68.2% HWCI, *black dashed lines*; SDs, *black dotted lines*), E-ETDRS (*red asterisks*), and FrACT (*green dotted lines*) methods as functions of testing time in the no-foil condition (a) and three foil conditions (b), with between-block variability correction, in the psychophysical experiment.
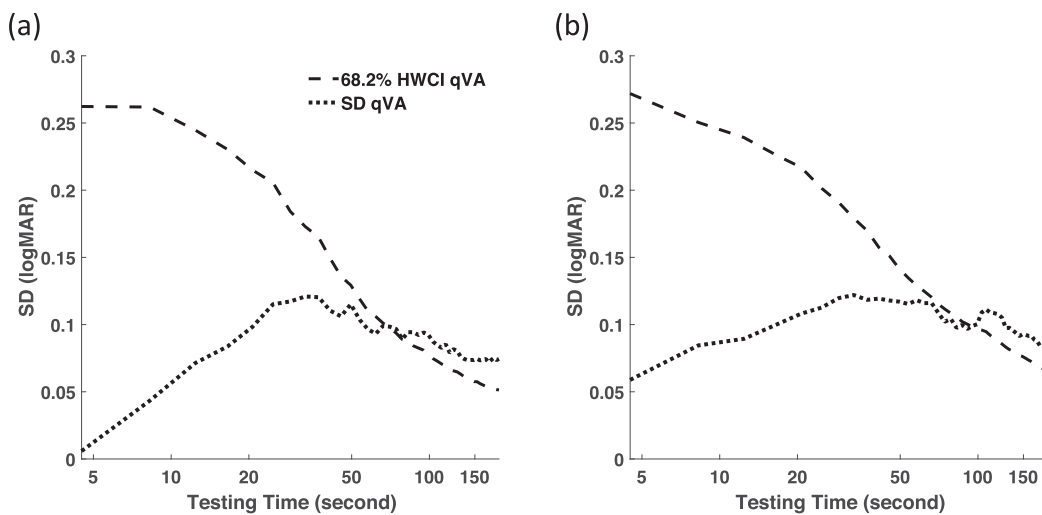

Figure 14. SDs of the estimated range from the qVA (68.2% HWCI, *black dashed lines*; SDs, *black dotted lines*) method as functions of testing times in the no-foil condition (a) and three foil conditions (b) in the psychophysical experiment.

Figures 11 and 12 and Table 3 show the precision of the VA threshold and VA range estimates as functions of number of tested optotypes. Figures 13 and 14 show the precision of the qVA estimates of threshold and

Table 4. Correlation Coefficients of VA Estimates from Three Psychophysical Methods

Method Pair		Correlation Coefficient	
		Data	Simulation
qVA	FrACT	0.9527	0.9783 ± 0.0025
	E-ETDRS	0.9541	0.9550 ± 0.0041
FrACT	E-ETDRS	0.9377	0.9476 ± 0.0058

range as functions of testing time. After 135 optotypes, the SD and 68.2% HWCI of the estimated VA threshold were 0.028 and 0.016 logMAR, respectively, in the no-foil condition, and 0.039 and 0.020 logMAR, respectively, in the three foil conditions. After 135 optotypes, the SDs and 68.2% HWCI of the estimated range were 0.074 and 0.051 logMAR, respectively, in the no-foil condition and 0.082 and 0.066 logMAR, respectively, in the three foil conditions.

Comparison of Estimated VA from the qVA, E-ETDRS, and FrACT Methods

We also compared the VA estimates from the qVA, E-ETDRS, and FrACT methods (Supplemen-
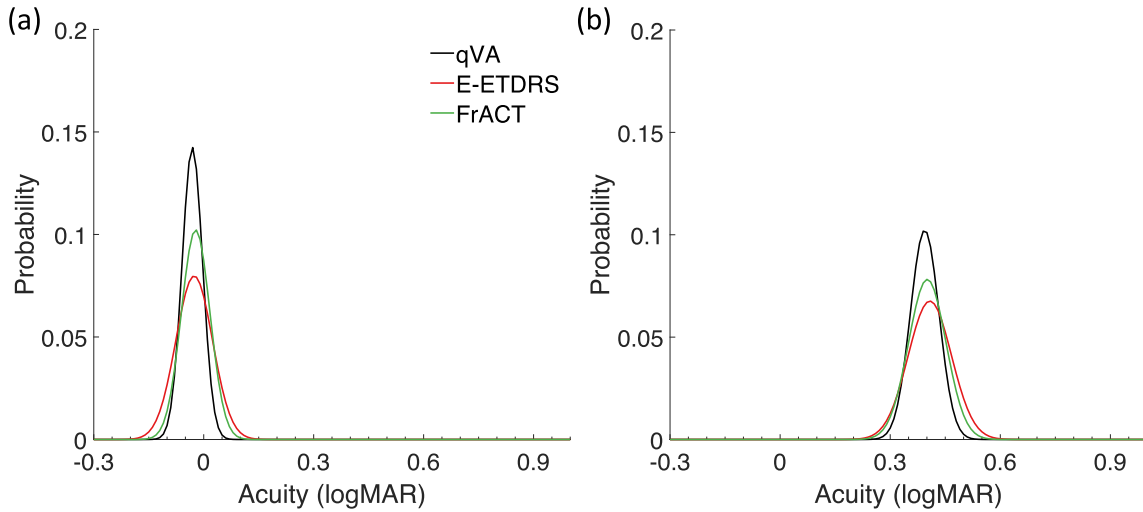
Figure 15. Mean distributions of the estimated VA threshold from the qVA (*black lines*), E-ETDRS (*red lines*), and FrACT (*green lines*) methods for the no-foil condition (a) and three foil conditions (b) across all the observers in the psychophysical experiment.

tary Fig. S6). Table 4 shows that the VA estimates from the three methods were all highly correlated, with values very close to their upper bounds based on simulations. In both the no-foil and three foil conditions (Fig. 15; Tables 3, 5), the qVA estimates exhibited higher precision (SD = 0.028 and 0.039 logMAR; FRP = 0.822 and 0.759) than the E-ETDRS (SD = 0.050 and 0.059 logMAR; FRP = 0.696 and 0.727) and FrACT (SD = 0.039 and 0.051 logMAR; FRP = 0.655 and 0.745) methods. The SDs of the estimated VA from the three methods were significantly different: $F(2, 156)$ = 5.695, P = 0.004. Post hoc Tukey test showed that qVA was significantly more precise than E-ETDRS (t = −3.319, P_{tukey} = 0.003) and marginally more precise than FrACT (t = −2.190, P_{tukey} = 0.076). At 95% specificity, the sensitivity values for detecting an estimated change of 0.15 logMAR were 96.5%, 77.6%, and 56.4% in the no-foil condition and were 77.6%, 54.8%, and 43.6% in the three foil conditions for the qVA, FrACT, and E-ETDRS methods, respectively (Table 6).

When the number of optotypes was matched, the qVA exhibited similar precision in estimating VA threshold (SD = 0.049 and 0.057 logMAR, FRP = 0.716 and 0.713 after 30 optotypes; 0.044 and 0.051 logMAR, FRP = 0.753 and 0.728 after 45 optotypes in the no-foil and three foil conditions) as the E-ETDRS (SD = 0.050 and 0.059 logMAR, FRP = 0.696 and 0.727 after 30 optotypes) and FrACT (SD = 0.039 and 0.051 logMAR, FRP = 0.655 and 0.745 after 45 optotypes) methods.

The qVA method, however, was the most efficient because of the shortest testing time (qVA, 41 and 62 seconds for 30 and 45 optotypes, respectively;

E-ETDRS, 66 seconds for 30 optotypes; FrACT, 112 seconds for 45 optotypes), and the average testing times per optotype were significantly different across the three methods, $F(2, 549)$ = 262.8, $P < 0.001$, with that of the qVA being significantly shorter than that of FrACT (t = −22.63, $P_{tukey} < 0.001$) and that of E-ETDRS (t = −15.43, $P_{tukey} < 0.001$). When the testing time was matched (the average testing times for 45 and 78 optotypes in the qVA were the same as those for 30 optotypes in the E-ETDRS and 45 optotypes in the FrACT tests, respectively), the SD of the estimated VA from qVA was much smaller than the SDs from the E-ETDRS and FrACT tests, except that it was the same as that of the FrACT test in the no-foil condition (Table 3). On average, the SD of the estimated VA from qVA was 35.7% less than that from the E-ETDRS and 24.3% less than that from the FrACT. These translate into 35.6% and 21.9% increases of sensitivity at 95% specificity (Table 6).

Table 5. FRP of VA Estimated from Three Psychophysical Methods

Method	Number of Optotypes	FRP	
		No Foil	Foil
qVA	30	0.716	0.713
	45	0.753	0.728
	135	0.822	0.759
E-ETDRS	30	0.696	0.727
FrACT	45	0.655	0.745

Table 6. Sensitivity at 95% Specificity to Detect an Estimated Change of 0.15 logMAR for Three Psychophysical Methods

Condition	Method	Number of Optotypes	SD	Sensitivity (%) at 95% Specificity
No foil	qVA	30	0.049	58.1
		45	0.044	67.4
		78	0.039	77.6
		135	0.028	96.6
	E-ETDRS	30	0.050	56.4
	FrACT	45	0.039	77.6
Three foil average	qVA	30	0.057	46.1
		45	0.051	54.8
		78	0.043	69.4
		135	0.039	77.6
	E-ETDRS	30	0.059	43.6
	FrACT	45	0.051	54.8

Discussion

In both computer simulations and a psychophysical experiment, we found that qVA estimates of the threshold and range of the VA behavioral function were accurate and precise. In the psychophysical experiment with 14 eyes and four Bangerter foil conditions, we found that qVA estimates of threshold and range were significantly correlated, and qVA threshold estimates were correlated with estimates from E-ETDRS and FrACT. Although more optotypes were needed for precise and accurate estimates of both VA threshold and range, the qVA required 38% and 45% less testing time to reach the same precision of the estimated VA threshold than the E-ETRDS and FrACT methods, respectively. The qVA can accurately, precisely, and efficiently quantify the full visual acuity behavioral function.

Comparison of VA Estimates

As discussed in the introduction, both the threshold and range of the VA behavioral function are necessary to fully characterize and interpret acuity scores and their changes. The qVA is the only method that can estimate both parameters that define the full behavioral function. To evaluate the qVA method, we compared its performance in estimating the VA threshold (while simultaneously estimating the range) with that of the E-ETDR and FrACT methods, which were developed to estimate only VA threshold.

We could not perform direct comparisons of the estimated VA thresholds from the three methods because (1) different stimulus configurations were used, and, more importantly, (2) they have different target threshold performance levels (in fact, not defined in E-ETDRS). On the other hand, the pairwise correlations among the estimated VA thresholds from the three methods in the psychophysical experiment suggest high consistencies. Although they are slightly lower than the upper bounds based on simulations (because only measurement errors but no observer variability were included in the simulations), the results provide sufficient validation for the VA threshold estimates from the qVA. The high correlations among the three methods indicated that the VA estimates from these methods tended to move in the same direction. A subject with a large VA estimate from one method was more likely to obtain a large VA estimate from another method, and vice versa. However, the high correlations did not indicate similar precision among the three methods (Table 3).

When the number of optotypes was matched between qVA and the other two methods, the SDs of the estimated VA from the three methods (and thus the sensitivity at 95% specificity) were similar in the psychophysical experiment. In the simulations, the SDs of the estimated VA from the qVA method were the smallest among the three tests when the number of optotypes was matched. The different results from the simulations and human psychophysical experiment may have resulted from the additional variability related to internal noise of the patient (such as random fluctuations in visual processing, attentional state, decision) and variability of foil alignment that was only present in the psychophysical experiment. In addition, that the precision of the estimated VA from the qVA increases with the number of optotypes is an

advantage of the qVA method. It allows clinicians to choose the number of tested optotypes based on their own target precision level.

The estimated VA thresholds from the qVA method were also more accurate and marginally more precise than those from the FrACT method and were more precise than those from the E-ETDRS method. Figures 15a and 15b illustrate the distributions of estimated VA thresholds from the three methods in the computer simulations. The qVA method generated unbiased VA threshold estimates with the highest precision. The bias of the estimated VA thresholds from the FrACT method is a result of the mismatch between the fixed slope used in the method and the true slopes of the simulated observers (Lu Z-L, et al. *IOVS*. 2019;60:ARVO E-Abstract 3908).

In addition, the SD (random measurement error) of the estimated VA threshold increases with the range of the VA behavioral function. Although the range and the SD of the estimated VA threshold can be correlated in a given measurement, they are different concepts. Whereas the range of the VA behavioral function is an intrinsic property of an observer that quantifies how performance changes as a function of optotype size, the SD of the estimated VA threshold quantifies the variability of a measurement and depends on the details of the measurement (e.g., instrument, number of trials). Furthermore, the random measurement error caused by binomial variability intrinsic in the task design (correct or incorrect response) and by internal noise of the patient (such as random fluctuations in visual processing, attentional state, decision) cannot be distinguished in the current study because both sources contribute to measurement error in every trial. The random measurement error of a patient with stable vision should also be distinguished from the true VA change of a patient whose visual disease progressed or was treated between two VA measurements.

Although the reanalysis of the E-ETDRS and FrACT data did improve the accuracy and precision of the VA estimates from the two methods to some extent (Tables 1, 2), the analysis algorithm of the qVA—the Bayesian procedure coupled with the full acuity behavioral psychometric functions—is only one advantage of the qVA method. The Bayesian active learning algorithm and high-density sampling of optotype size in the qVA are critical for optimizing information gain in each trial and obtaining accurate and precise VA threshold and range estimates. The ETDRS and FrACT methods are not optimal for measuring the full acuity behavioral function (see below).

Range or Slope of the Acuity Psychometric Function

The qVA method was developed to measure both VA threshold and range (or slope) of the acuity psychometric function (Patent No. US 10758120B2; Lesmes LA. *IOVS*. 2018;59:ARVO E-Abstract 1073).[41] We found a significant main effect of foil condition, $F(3, 18) = 9.111$, $P < 0.001$, for the steepness of the psychometric function (quantified by the range parameter) and a significant correlation between the estimated VA threshold and range ($r = 0.412$, $P < 0.001$), consistent with Carkeet et al.,[27] who found that the steepness of the acuity psychometric function depended on the level of blur. Therefore, slope might be a potential endpoint in clinical vision with change measurable by the qVA method.

On the other hand, the E-ETDRS and FrACT methods, like all other VA tests, provide only a point estimate of the VA threshold. Reanalysis of the data from the E-ETDRS and FrACT methods in the simulations with the Bayesian procedure in the qVA method showed that the data from them did not provide sufficient information to constrain range; its 68.2% HWCI was at least twice of its SD from repeated runs. Figure 4 illustrates that point with one simulated run of simulated Observer 1. In the qVA method (Figs. 4a, 4d), a wide range of optotype sizes is sampled to obtain an accurate and precise estimate of both VA threshold and range of the VA behavioral function. In the E-ETDRS method (Figs. 4b, 4e), although the screening phase covers a large range of optotype sizes, an insufficient number of optotypes were tested to constrain the range estimate. In the FrACT method (Figs. 4c, 4f), the sizes of the test stimuli were mostly concentrated near the 55% correct VA threshold after about 10 optotypes and were insufficient to constrain the range estimate. As a consequence, the estimated VA range from the simulated FrACT data exhibited the largest bias (86% and 233% greater than those from the qVA for Observers 1 and 2) (Table 1).

To further demonstrate the importance of the full acuity function in estimating VA change, we simulated visual acuity behavior change for Observer 1. In this simulation, the observer's VA threshold increased by 0.15 logMAR and the VA range increased by 0.3 logMAR. We found the average estimated VA threshold changes were 0.15 logMAR, 0.09 logMAR, and 0.09 logMAR from the qVA, E-ETDRS, and FrACT methods, respectively. Without explicitly considering the range of the acuity psychometric function, both E-ETDRS and FrACT underestimated the VA change by 0.06 logMAR (3 ETDRS letters).

Multi-Optotype Psychometric Function and Joint Distributions

Although the qVA method employs the same Bayesian adaptive framework as other methods,[42–61] such as QUEST+,[54] the qVA method is tailored to the specifics involved in the visual acuity measurement. Specifically, it models the full acuity behavior via composite multiple-optotype psychometric functions derived from a single-optotype d' acuity psychometric function to correct guessing behavior for acuity tests that often involve optotype sampling without replacement. To our best knowledge, this composite multiple-optotype model is the first that has been incorporated into an adaptive method to measure full VA behavior.

Most adaptive testing algorithms[46–48,52,54,55] are sensitive to lapse, especially if it occurs early in the measurement. However, the qVA method uses an adaptive testing algorithm that is robust to lapse, based on many years of development on Bayesian adaptive tests by our group.[41–45,49–51,53,56–61] Most adaptive tests are sensitive to lapse because they assume that the observer's behavior obeys the underlying psychometric model and selects test stimulus to maximize the expected information gain in the next trial. However, because lapse is not considered, the model is incorrect if lapse occurs; thus, maximizing the expected information gain in the next trial would lead the algorithm astray. In our development, perturbation was added to the expected information gain to "shake" the tests so that they would not be trapped by lapse.[43]

In addition, the joint posterior distribution of VA threshold and range from the qVA method provides rich information for further statistical analyses. For example, the 68.2% HWCI of the joint posterior distribution quantifies the within-run variability of the estimated VA threshold and range, which is not available from traditional tests with only point estimates. The joint posterior distributions can also be incorporated into hierarchical models to improve statistical power in hypothesis testing at both individual and group levels (Lu Z-L, et al. *IOVS*. 2020;61:ARVO E-Abstract 4615; Zhao Y, et al. *IOVS*. 2020;61:ARVO E-Abstract 4616).

Future Development

The introduction of the composite multiple-optotype psychometric functions based on a single-optotype d' acuity psychometric function is one of the key innovations of the qVA method. It is used to accommodate different tasks and chart designs. d' is a task-invariant measure of an observer's ability to distinguish different stimuli along a selected stimulus dimension, whereas percent correct quantifies only the empirical performance in a task and depends on the details of the chart design. Although in this study we tested adults with only a 10AFC task using high-contrast Sloan letters, different optotypes (e.g., pictogram optotypes[78,79]), tasks (e.g., Yes–No task[51]), and/or test conditions (e.g., test distance), tailored to different ages, cognitive abilities, VA ranges, and/or disease characteristics, could be incorporated into the qVA method, especially given that some new chart designs have been shown to improve disease diagnosis, management, and/or treatment evaluations, such as the low luminance chart[80] or high-pass letter chart[15,81,82] for age-related macular degeneration patients and low contrast chart for multiple sclerosis.[11,83–85]

In this study, weakly informative prior distributions were used for each parameter based on pilot experiments. Because a strong incorrect prior would take extra trials for the algorithm to find the correct estimate,[86] a weakly informative prior is a better option when very little information is available. It would avoid the extra trials needed to obtain a reliable estimate with a strong but incorrect prior. On the other hand, the accuracy, precision, and efficiency of the qVA method can be improved by using more informative priors.[86,87] In addition, the correlation between VA and range could be incorporated into the prior to further improve the efficiency of qVA estimation. Prior knowledge can also be integrated into the estimation procedure in the form of hierarchical modeling.[86,87] In addition, increasing the number of optotypes on a row (e.g., from 3 to 5) in a trial increases the information collected in each trial so the efficiency of the qVA estimation could be further improved. Because estimating the VA range requires more optotypes than estimating the VA threshold, increased efficiency is more critical in estimating range. The "I don't know" response option usually presented in unforced-choice tasks[88] was introduced to improve patient experience in quick contrast-sensitivity function tests.[44,50] We re-simulated the qVA method for simulated Observer 1 with the "I don't know" response based on its frequency in psychophysical experiment. We found that the response option introduced a relatively small 0.022 logMAR bias to the estimated VA threshold. We will incorporate the model of unforced-choice tasks[88] into the qVA method to reduce the bias in the future.

To the best of our knowledge, this is the first study to validate the novel qVA method in both simulations and psychophysical experiment. Although we used only seven subjects in the study, we collected a massive amount of data from each of them (6720 optotypes over four foil conditions). Nevertheless, the results are based on a relatively small number of subjects with

normal or corrected-to-normal vision and high education level. More subjects and different patient populations are necessary to further evaluate the qVA method in the future.

Conclusions

In both computer simulations and a psychophysical experiment, we showed that the qVA method can accurately, precisely, and efficiently quantify the full VA acuity behavior using an acuity psychometric function with two parameters: a VA threshold at $d' = 2$ performance level and range. The concurrent measurement of VA and slope did not compromise the VA measurement, as supported by the relatively high precision and efficiency of the estimated VA threshold from the qVA method.

Acknowledgments

References

1. Kniestedt C, Stamper RL. Visual acuity and its measurement. *Ophthalmol Clin North Am*. 2003;16(2):155–170.
2. Levenson JH, Kozarsky A. Visual acuity. In: Walker HK, Hall WD, Hurst JW, eds. *Clinical Methods: The History, Physical, and Laboratory Examinations*. 3rd ed. Boston, MA: Butterworths; 1990.
3. Clare G, Pitts JA, Edgington K, Allan BD. From beach lifeguard to astronaut: occupational vision standards and the implications of refractive surgery. *Br J Ophthalmol*. 2010;94(4):400–405.
4. World Para Alpine Skiing. Classification in para alpine skiing. Available at: https://www.paralympic.org/alpine-skiing/classification. Accessed December 17, 2020.
5. Rahi JS, Cumberland PM, Peckham CS. Does amblyopia affect educational, health, and social outcomes? Findings from 1958 British birth cohort. *Br Med J*. 2006;332(7545):820–824.
6. Kiel AW, Butler T, Alwitry A. Visual acuity and legal visual requirement to drive a passenger vehicle. *Eye*. 2003;17(5):579–582.
7. Lennie P, Van Hemel SB, eds. *Visual Impairments: Determining Eligibility for Social Security Benefits*. Washington, DC: National Academy Press; 2002.
8. NICE. *Age-Related Macular Degeneration—NICE Guideline (NG82)*. Manchester, UK: National Institute for Health and Care Excellence; 2018.
9. Anstice NS, Thompson B. The measurement of visual acuity in children: an evidence-based update. *Clin Exp Optom*. 2014;97(1):3–11.
10. Bailey IL, Lovie JE. New design principles for visual acuity letter charts. *Am J Optom Physiol Opt*. 1976;53(11):740–745.
11. Balcer LJ, Baier ML, Pelak VS, et al. New low-contrast vision charts: reliability and test characteristics in patients with multiple sclerosis. *Mult Scler*. 2000;6(3):163–171.
12. Ferris FL 3rd, Kassoff A, Bresnick GH, Bailey I. New visual acuity charts for clinical research. *Am J Ophthalmol*. 1982;94(1):91–96.
13. Plainis S, Tzatzala P, Orphanos Y, Tsilimbaris MK. A modified ETDRS visual acuity chart for European-wide use. *Optom Vis Sci*. 2007;84(7):647–653.
14. Rosser DA, Laidlaw DA, Murdoch IE. The development of a "reduced logMAR" visual acuity chart for use in routine clinical practice. *Br J Ophthalmol*. 2001;85(4):432–436.
15. Shah N, Dakin SC, Whitaker HL, Anderson RS. Effect of scoring and termination rules on test–retest variability of a novel high-pass letter acuity chart. *Invest Ophthalmol Vis Sci*. 2014;55(3):1386–1392.
16. Snellen H. *Optotypi ad Visum Determinandum*. Utrecht: P.W. van de Weijer; 1862.
17. Beck RW, Moke PS, Turpin AH, et al. A computerized method of visual acuity testing: adaptation of the early treatment of diabetic retinopathy study testing protocol. *Am J Ophthalmol*. 2003;135(2):194–205.
18. Bokinni Y, Shah N, Maguire O, Laidlaw DAH. Performance of a computerised visual acuity measurement device in subjects with age-related macular degeneration: comparison with gold

standard ETDRS chart measurements. *Eye*. 2015;29(8):1085–1091.

19. Laidlaw DAH, Tailor V, Shah N, Atamian S, Harcourt C. Validation of a computerised logMAR visual acuity measurement system (COMPlog): comparison with ETDRS and the electronic ETDRS testing algorithm in adults and amblyopic children. *Br J Ophthalmol*. 2008;92(2):241–244.

20. Ehrmann K, Fedtke C, Radic A. Assessment of computer generated vision charts. *Contact Lens Anterior Eye*. 2009;32(3):133–140.

21. Gonzalez EG, Tarita-Nistor L, Markowitz SN, Steinbach MJ. Computer-based test to measure optimal visual acuity in age-related macular degeneration. *Invest Ophthalmol Vis Sci*. 2007;48(10):4838–4845.

22. Perera C, Chakrabarti R, Islam FMA, Crowston J. The Eye Phone Study: reliability and accuracy of assessing Snellen visual acuity using smartphone technology. *Eye (Lond)*. 2015;29(7):888–894.

23. Schlenker MB, Christakis TJ, Braga-Mele RM. Comparing a traditional single optotype visual acuity test with a computer-based visual acuity test for childhood amblyopia vision screening: a pilot study. *Can J Ophthalmol*. 2010;45(4):368–374.

24. Sreelatha OK, Ramesh SV, Jose J, Devassy M, Srinivasan K. Virtually controlled computerised visual acuity screening in a multilingual Indian population. *Rural Remote Health*. 2014;14(3):2908.

25. Srinivasan K, Ramesh SV, Babu N, Sanker N, Ray A, Karuna SM. Efficacy of a remote based computerised visual acuity measurement. *Br J Ophthalmol*. 2012;96(7):987–990.

26. Bach M. The Freiburg Visual Acuity Test-variability unchanged by post-hoc re-analysis. *Graefes Arch Clin Exp Ophthalmol*. 2007;245(7):965–971.

27. Carkeet A, Lee L, Kerr JR, Keung MM. The slope of the psychometric function for Bailey-Lovie letter charts: defocus effects and implications for modeling letter-by-letter scores. *Optom Vis Sci*. 2001;78(2):113–121.

28. Carkeet A, Bailey IL. Slope of psychometric functions and termination rule analysis for low contrast acuity charts. *Ophthalmic Physiol Opt*. 2017;37(2):118–127.

29. Hazel CA, Elliott DB. The dependency of logMAR visual acuity measurements on chart design and scoring rule. *Optom Vis Sci*. 2002;79(12):788–792.

30. Bach M. The Freiburg Visual Acuity test–automatic measurement of visual acuity. *Optom Vis Sci*. 1996;73(1):49–53.

31. Carkeet A. Modeling logMAR visual acuity scores: effects of termination rules and alternative forced-choice options. *Optom Vis Sci*. 2001;78(7):529–538.

32. Gibson R, Sanderson H. Observer variation in ophthalmology. *Br J Ophthalmol*. 1980;64(6):457–460.

33. Laidlaw DAH, Abbott A, Rosser DA. Development of a clinically feasible logMAR alternative to the Snellen chart: performance of the "compact reduced logMAR" visual acuity chart in amblyopic children. *Br J Ophthalmol*. 2003;87(10):1232–1234.

34. Lovie-Kitchin JE. Is it time to confine Snellen charts to the annals of history? *Ophthalmic Physiol Opt*. 2015;35(6):631–636.

35. McGraw P, Winn B, Whitaker D. Reliability of the Snellen chart. *Br Med J*. 1995;310(6993):1481–1482.

36. Arditi A, Cagenello R. On the statistical reliability of letter-chart visual acuity measurements. *Invest Ophthalmol Vis Sci*. 1993;34(1):120–129.

37. Elliott DB, Sheridan M. The use of accurate visual-acuity measurements in clinical anti-cataract formulation trials. *Ophthalmic Physiol Opt*. 1988;8(4):397–401.

38. Vanden Bosch ME, Wall M. Visual acuity scored by the letter-by-letter or probit methods has lower retest variability than the line assignment method. *Eye (Lond)*. 1997;11(pt 3):411–417.

39. Rosser DA, Murdoch IE, Fitzke FW, Laidlaw DAH. Improving on ETDRS acuities: design and results for a computerised thresholding device. *Eye (Lond)*. 2003;17(6):701–706.

40. Shah N, Laidlaw DAH, Shah SP, Sivasubramaniam S, Bunce C, Cousens S. Computerized repeating and averaging improve the test-retest variability of ETDRS visual acuity measurements: implications for sensitivity and specificity. *Invest Ophthalmol Vis Sci*. 2011;52(13):9397–9402.

41. Lesmes LA, Dorr M. Active learning for visual acuity testing. In: *Proceedings of the 2nd International Conference on Applications of Intelligent Systems, APPIS '19*. New York, NY: Association for Computing Machinery; 2019:1–6.

42. Baek J, Lesmes LA, Lu Z-L. qPR: an adaptive partial-report procedure based on Bayesian inference. *J Vis*. 2016;16(10):1–23.

43. Hou F, Lesmes L, Bex P, Dorr M, Lu Z-L. Using 10AFC to further improve the efficiency of the quick CSF method. *J Vis*. 2015; 15(9):1–18.

44. Hou F, Lesmes LA, Kim W, et al. Evaluating the performance of the quick CSF method in detecting contrast sensitivity function changes. *J Vis*. 2016;16(6):1–19.

45. Hou F, Zhao Y, Lesmes LA, Bex P, Yu D, Lu Z-L. Bayesian adaptive assessment of the reading function for vision: the qReading method. *J Vis*. 2018;18(9):1–16.

46. King-Smith PE, Grigsby SS, Vingrys AJ, Benes SC, Supowit A. Efficient and unbiased modifications of the QUEST threshold method: theory, simulations, experimental evaluation and practical implementation. *Vision Res*. 1994;34(7):885–912.

47. Kontsevich LL, Tyler CW. Bayesian adaptive estimation of psychometric slope and threshold. *Vision Res*. 1999;39(16):2729–2737.

48. Kujala JV, Lukka TJ. Bayesian adaptive estimation: the next dimension. *J Math Psychol*. 2006;50(4):369–389.

49. Lesmes LA, Jeon S-T, Lu Z-L, Dosher BA. Bayesian adaptive estimation of threshold versus contrast external noise functions: the quick TvC method. *Vision Res*. 2006;46(19):3160–3176.

50. Lesmes LA, Lu Z-L, Baek J, Albright TD. Bayesian adaptive estimation of the contrast sensitivity function: the quick CSF method. *J Vis*. 2010;10(3):1–21.

51. Lesmes LA, Lu Z-L, Baek J, Tran N, Dosher BA, Albright TD. Developing Bayesian adaptive methods for estimating sensitivity thresholds (d') in Yes-No and forced-choice tasks. *Front Psychol*. 2015;6:1070.

52. Prins N. The psi-marginal adaptive method: how to give nuisance parameters the attention they deserve (no more, no less). *J Vis*. 2013;13(7):1–17.

53. Shepard TG, Hou F, Bex PJ, Lesmes LA, Lu Z-L, Yu D. Assessing reading performance in the periphery with a Bayesian adaptive approach: the qReading method. *J Vis*. 2019;19(5):1–14.

54. Watson AB. QUEST+: a general multidimensional Bayesian adaptive psychometric method. *J Vis*. 2017;17(3):1–27.

55. Watson AB, Pelli DG. Quest: a Bayesian adaptive psychometric method. *Percept Psychophys*. 1983;33(2):113–120.

56. Xu P, Lesmes LA, Yu D, Lu Z-L. A novel Bayesian adaptive method for mapping the visual field. *J Vis*. 2019;19(14):1–32.

57. Xu P, Lesmes LA, Yu D, Lu Z-L. Mapping the contrast sensitivity of the visual field with Bayesian adaptive qVFM. *Front Neurosci*. 2020;14:665.

58. Zhang P, Zhao Y, Dosher BA, Lu Z-L. Assessing the detailed time course of perceptual sensitivity change in perceptual learning. *J Vis*. 2019;19(5):1–19.

59. Zhang P, Zhao Y, Dosher BA, Lu Z-L. Evaluating the performance of the staircase and quick Change Detection methods in measuring perceptual learning. *J Vis*. 2019;19(7):1–25.

60. Zhao Y, Lesmes LA, Lu Z-L. Efficient assessment of the time course of perceptual sensitivity change. *Vision Res*. 2019;154:21–43.

61. Lu Z-L, Dosher BA. *Visual Psychophysics: From Laboratory to Theory*. Cambridge, MA: MIT Press; 2013.

62. Jensen AL, Kjelgaard-Hansen M. Method comparison in the clinical laboratory. *Vet Clin Pathol*. 2006;35(3):276–286.

63. Bland J, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1(8476):307–310.

64. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res*. 1999;8(2):135–160.

65. Bailey I, Bullimore M, Raasch T, Taylor H. Clinical grading and the effects of scaling. *Invest Ophthalmol Vis Sci*. 1991;32(2):422–432.

66. Dorr M, Lesmes LA, Elze T, Wang H, Lu Z-L, Bex PJ. Evaluation of the precision of contrast sensitivity function assessment on a tablet device. *Sci Rep*. 2017;7:46706.

67. Clayton D, Hills M. *Statistical Models in Epidemiology*. Oxford, UK: Oxford University Press; 1993.

68. Edwards W, Lindman H, Savage LJ. Bayesian statistical inference for psychological research. *Psychol Rev*. 1963;70(3):193–242.

69. Rosser DA, Cousens SN, Murdoch IE, Fitzke FW, Laidlaw DAH. How sensitive to clinical change are ETDRS logMAR visual acuity measurements? *Invest Ophthalmol Vis Sci*. 2003;44(8):3278–3281.

70. Cousens SN, Rosser DA, Murdoch IE, Laidlaw DA. A simple model to predict the sensitivity to change of visual acuity measurements. *Optom Vis Sci*. 2004;81(9):673–677.

71. Csaky K, Ferris F, Chew EY, Nair P, Cheetham JK, Duncan JL. Report from the NEI/FDA Endpoints Workshop on Age-Related Macular Degeneration and Inherited Retinal Diseases. *Invest Ophthalmol Vis Sci*. 2017;58(9):3456–3463.

72. Odell NV, Leske DA, Hatt SR, Adams WE, Holmes JM. The effect of Bangerter filters on optotype acuity, Vernier acuity, and contrast sensitivity. *J AAPOS*. 2008;12(6):555–559.

73. Kleiner M, Brainard D, Pelli D. What's new in Psychtoolbox-3? *Perception*. 2007;36:1–16.

74. Ricci F, Cedrone C, Cerulli L. Standardized measurement of visual acuity. *Ophthalmic Epidemiol*. 1998;5(1):41–53.

75. Bach M. Anti-aliasing and dithering in the "Freiburg Visual Acuity Test." *Spatial Vis*. 1997;11(1):85–89.

76. Pérez GM, Archer SM, Artal P. Optical characterization of Bangerter foils. *Invest Ophthalmol Vis Sci*. 2010;51(1):609–613.

77. Dorr M, Elze T, Wang H, Lu Z-L, Bex PJ, Lesmes LA. New precision metrics for contrast sensitivity testing. *IEEE J Biomed Health Inform*. 2018;22(3):919–925.

78. Hyvarinen L, Nasanen R, Laurinen P. New visual acuity test for pre-school children. *Acta Ophthalmol (Copenh)*. 1980;58(4):507–511.

79. Hamm LM, Yeoman JP, Anstice N, Dakin SC. The Auckland Optotypes: an open-access pictogram set for measuring recognition acuity. *J Vis*. 2018;18(3):1–15.

80. Chandramohan A, Stinnett SS, Petrowski JT, et al. Visual function measures in early and intermediate age-related macular degeneration. *Retina*. 2016;36(5):1021–1031.

81. Anderson RS. Improving ophthalmic diagnosis in the clinic using the Moorfields Acuity Chart. *Expert Rev Ophthalmol*. 2017;12(6):433–435.

82. Shah N, Dakin SC, Dobinson S, Tufail A, Egan CA, Anderson RS. Visual acuity loss in patients with age-related macular degeneration measured using a novel high-pass letter chart. *Br J Ophthalmol*. 2016;100(10):1346–1352.

83. Balcer LJ, Galetta SL, Polman CH, et al. Low-contrast acuity measures visual improvement in phase 3 trial of natalizumab in relapsing MS. *J Neurol Sci*. 2012;318(1–2):119–124.

84. Balcer LJ, Raynowska J, Nolan R, et al. Validity of low-contrast letter acuity as a visual performance outcome measure for multiple sclerosis. *Mult Scler J*. 2017;23(5):734–747.

85. Mowry EM, Loguidice MJ, Daniels AB, et al. Vision related quality of life in multiple sclerosis: correlation with new measures of low and high contrast letter acuity. *J Neurol Neurosurg Psychiatry*. 2009;80(7):767–772.

86. Gu H, Kim W, Hou F, et al. A hierarchical Bayesian approach to adaptive vision testing: a case study with the contrast sensitivity function. *J Vis*. 2016;16(6):1–17.

87. Kim W, Pitt MA, Lu Z-L, Steyvers M, Myung JI. A hierarchical adaptive approach to optimal experimental design. *Neural Comput*. 2014;26(11):2465–2492.

88. Kaernbach C. Adaptive threshold estimation with unforced-choice tasks. *Percept Psychophys*. 2001;63(8):1377–1388.