AIM (Angular Indication Measurement) paradigm: Adaptive, self-administered, and generalizable vision assessment used to measure visual acuity

Jan Skerswetat * [1], Jingyi He[1], Jay Bijesh Shah[1], Nicolas Aycardi[1], & Peter J. Bex [1]

[1] Department of Psychology, Northeastern University, 360 Huntington Ave, Boston, Massachusetts, 02115, United States of America

Corresponding Author *
Email address: j.skerswetat@northeastern.edu

Abstract

Purpose: Visual acuity (VA) is a critical clinical endpoint, however tests typically require trained staff, are fixed in their range, and provide only limited information about individual performance. This study introduces the Angular Indication Measurement and applies it to VA (AIM-VA). Agreement, repeatability, and test duration of AIM-VA were compared to standard clinical tests.

Methods: 29 and 22 adults took part in two separate experiments. 4*4 charts with Landolt-C optotypes of varying sizes and random gap-orientations were presented at 4m viewing distance. Optotype size was response-adaptively updated after each chart. Participants indicated the perceived gap location by clicking on the corresponding location on a surrounding response-ring and VA was estimated from the report errors as a function of target size. First, we compared AIM-VAs with that of ETDRS and FrACT. Second, we compared repeatability over two tests of AIM-VA and ETDRS. AIM-VA-specific psychometric functions generated a new parameter called range of stimulus detectability improvement (ROSDI) that represents the range over which VA varies with target size.

Results: There were no significant differences among VA estimates from AIM, ETDRS and FrACT. AIM-VA's test-retest repeatability was comparable to that of ETDRS (95% Confidence Intervals in logMAR: AIM-VA [0.18 / -0.15] when using 3 adaptive steps and ETDRS [0.18 / -0.11]). Median ROSDI was 0.14logMAR ±0.03. Test duration was not significantly different for AIM-VA and ETDRS, the latter being significantly faster than FrACT.

Conclusions: AIM-VA estimates of VA were equivalent to FrACT and ETDRS. AIM-VA can be estimated in approximately the same time as ETDRS and with comparable test-retest reliability. AIM-VA is an intuitive, self-administered, adaptive, and personalized method that provides reliable and repeatable estimates of VA and additional performance outcomes. AIM's orientation judgement approach is generalizable to interrogate many other visual functions, e.g., contrast, color, motion, stereo-vision.

Keywords: Visual acuity, psychophysical methods, response-adaptive, self-administered, platform technology

Introduction

Visual function assessment is an essential task for eye care clinicians. One of the major challenges for eyecare professionals worldwide is refractive errors as estimates suggest that approximately 2.6 billion people are affected by myopia and 1.8 billion people by presbyopia (1). Visual acuity (VA) is arguably the key endpoint in current clinical practice to assess central vision as it is sensitive to changes of the refraction of the eye (2), but also changes of central retinal (3) and ocular media structures (4), changes along the visual pathway (5), and atypical visual-cortical structures (6). Since the first known reports for VA tests by Tobias Mayer 1757 (7) and Heinrich Küchler in 1843 (8), various techniques and tasks have evolved to measure VA; however, all of those tests examine the observer's ability to discern different structural details from each other by subjective report, for recent review see (7).

In practice VA assessment typically requires a participant to identify letter or number optotypes in 4 alternative forced choice (AFC) letter charts such as HOTV (9), 10AFC letter charts such as Early Treatment of Diabetic Retinopathy Study (ETDRS) (10, 11) or 9AFC letter charts such as Snellen (12). Observers who are non-literate or unfamiliar with the Roman alphabet may be asked to identify the orientation of the gap of tumbling E (13) or Landolt C (14) optotypes in 4 or 8 AFC tasks; to match child-friendly optotypes such as Lea symbols (15) or Auckland Optotypes (16) in 10AFC tasks, or to report the offset of misaligned line stimuli in 2AFC tasks (17). For participants such as babies who unable to generate a verbal or manual report, preferential looking behavior can be converted to a 2AFC task (18).

Typically, optotypes are presented in printed or computer-generated charts in which the layout, size and spacing of optotypes is organized with monotonically increasing difficulty. The number and spacing of letters can vary (12) or may be logarithmic (19) following the Weber-Fechner law (20). Various rules have been developed to score the correct and incorrect responses of the participant, including letter scoring, in which VA is related to the total number of letters correctly identified; line scoring, in which VA is defined by the optotype size where a specified proportion of letters are correctly reported; and stopping criteria, all of which affect the estimate of VA and test-retest repeatability (21–23).

The transition to computer-administered tests avoids memorization artifacts of printed charts and allows the implementation of adaptive psychophysical testing principles (24, 25) and automated scoring eliminating inter-examiner differences (24, 26). More recently, numerous web-browser- and app-based VA tests have entered the field to monitor VA remotely (25, 27, 28), thereby eliminating the need for a clinic visit and personnel to administer the test. However, current tests are not regulated (29) and poor quality or unreliable tests may pose a clinical risk to consumers (30).

To address all of these issues, we introduce the Angular Indication Measurement (AIM) method, which addresses the shortcomings of current tests described above. AIM is a computer-based, generalizable, self-administered, response-adaptive platform. AIM measures multiple visual function endpoints by adapting different stimulus and response formats to an orientation judgment. In each case stimuli are presented at a range of personalized stimulus intensities and random orientations (acuity, color, contrast, stereoscopic disparity etc.) or directions (motion). The observer's task is to report the perceived orientation or direction of the stimulus, and the difference between the actual and reported stimulus orientation/direction provides a continuous error score as a function of stimulus intensity (31).

The current study has three major objectives. First, we introduce AIM and describe its features and outcome measures. Second, we applied AIM to interrogate VA, namely AIM-VA, and compare it with two standard VA tests, the retro-luminant ETDRS chart and the computer-based FrACT for testing duration and estimated VA. Third, we investigated the test-re-test reliability of AIM-VA and compared it with ETDRS results. Additionally, we will describe novel AIM-VA outcomes and report their summary statistics for the cohort tested in this study.

Methods

Participants were recruited from the Translational vision laboratory and from the undergraduate population at Northeastern University, Boston. Undergraduate students received course credit towards the completion of their Introductory Psychology course in exchange for their participation. Written and verbal information about the project were provided in advance to the participants and they gave informed consent via online form before taking part. The internal reviewing board of Northeastern University approved the study and was in line with the ethical principles of the Helsinki declaration of 1975.

Two separate experiments were conducted, and their methods are reported below. The first experiment compared AIM-VA tom ETDRS and FrACT in terms of VA and testing duration. The second experiment investigated AIM-VA's test-retest reliability.

Participants

Experiment 1: 29 adults. Stereoacuities and central interocular suppression were measured using the RanDot test at 41cm and a Worth 4-Dot test at a viewing distance of 1m was carried out to test for central interocular suppression. The Miles eye dominance test was carried out to determine the eye dominance. We also asked whether the participant was left or right-handed and whether they wore glasses, contact lenses, or no correction.

Experiment 2: 23 adults. An objective refraction was performed using autorefractometer (AR(K)7610 Auto Refracto-Keratometer) and prescription of glasses if worn, were measured using an auto-focimeter (Xinyuan, Auto Lensmeter). We also asked for handedness, used the hole-in-a-card method to determine eye dominance, and whether they wore glasses, contact lenses, or no correction.
Control experiment: An additional control experiment was performed to determine the effect of the number of adaptive steps on test-retest repeatability, 12 participants were recruited, including a subset of 7 participants from Experiment two and 5 additional participants (3 with and 2 without correction).

Equipment

Stimuli were presented on a HP (USA) All-in-One computer and monitor system (Model: 32-a1027c Rfrbd PC) with a framerate of 60 Hz with a pixel resolution of [3840 / 2160], that subtended 10° at a 4m viewing distance with a resolution of 384 pixels per degree. Stimuli were generated using Matlab (MathWorks, R2021a) software in combination with *The Psychophysics Toolbox* (32, 33). An HP mouse was used as indication device. FrACT (version 3.10.5) software (24) was applied on the same monitor. We used a retro-luminant ETRDS chart and illuminator (Lighthouse, USA, second edition). The monitor was gamma-corrected using datacolor Spyder (Switzerland) photometer. We used a retro-luminant ETRDS chart (Lighthouse, USA, second edition) and measured the background luminance with a LS-110 (Konica Minolta Sensing Americas) photometer.

AIM principal

AIM is a generalizable psychophysical paradigm in which one or more stimuli is presented at a random orientation, and the observer's task is to indicate on a surrounding response ring the perceived orientation of the target, using a mouse or touch response. The difference between the true and reported orientation provides a continuous error score for each stimulus test level/intensity. Across different tests, stimulus intensity, can refer to contrasts and spatial frequencies (34), cone-isolated colors (35), stereoscopic disparity (31) etc. In the current study, we varied optotype size to interrogate VAs.

We used Landolt C optotypes, which are recognized as the laboratory standard by the International Council of Ophthalmology (31) because they avoid several problems with numbers and letters which differ in their size, critical identifying details and recognizability, in addition to literacy requirements described above. Landolt C optotype tests have the advantage that the physical and spectral properties of optotypes do not vary within a chart except by the gap orientation and do not vary in meaning, unlike object optotypes, such as letters, hence the acuity determine with

such stimuli is a more unbiased estimate of an observer's VA. Following International Council of Ophthalmology (31) recommendations, Landolt Cs were defined as interrupted circles whose stroke width and gap width are one-fifth of its outer diameter.
We followed and presented a range of optotypes with logarithmically spaced sizes. C optotypes were presented on a series of 2 grids, each 4x4 cells of 1.3° cells surrounded by a 0.1° thick indication ring. The background luminance and stimulus luminance were adjusted to match that of the ETDRS. The pixel (square) size was 0.0026° (384 pixel/1°). The C' gap width was kept at 20% of the line width. The inter-cell gap distance was 0.1°. The surrounding rings had a grey hue and changed color to blue to visualize an indication was made as well as the appearance of a black arc that indicated the gap orientation (Figure 1 A).

The stimulus sizes varied from difficult to easy, with log-step sizes starting from -0.3 to 0.8logMAR on the 1st chart, following the design principle of the ETDRS chart (19). The range of sizes on the second and subsequent charts were in equal log steps between the latest threshold size estimate ±2σ from the fit to the results of previous charts. The orientation of the optotype in each cell was drawn from a uniform random distribution and the different optotype sizes were randomly assigned to each cell. Participants indicated the perceived orientation of all 16 stimuli by clicking on each ring (Figure 1 A), guessing if unsure. Participants could repeatedly correct their choice by clicking again on any response ring and proceeded to the next chart by clicking on a 'Next' button that became available only once responses had been made to all 16 stimuli.

The response errors for each optotype size were converted to 8AFC responses, where absolute orientation errors less than 22.5° was scored as correct response, and errors greater than 22.5° were scored as incorrect responses. A cumulative gaussian psychometric function was fit to the 8AFC data, from which VA threshold ±2σ was used to select the optotype sizes for the next chart.

At the end of the final chart, a cumulative gaussian function was fit to the error report data as a function of optotype size from all charts, defined as:

$$\tau_\theta(s) = \left(\theta_i + \left(\theta_{guess} - \theta_i\right)\right)\left(0.5 + 0.5 * erf\left(\frac{s-\tau}{\sqrt{2}\gamma}\right)\right) \qquad (1)$$

Where θ_{guess} is orientation error for a guess response at 90° for a stimulus with 360° range e.g., Landolt C (or 45° for a stimulus with 180° range e.g. grating-type stimuli), *s* is signal intensity (here stimulus size), *τ* is a sensitivity threshold, θ_i is internal angular uncertainty (*noise* – the angular report error for the most visible stimulus) , and γ is the slope of the function. Figure 1C shows example data for one observer where the x axis shows optotype size in logMAR units, the y axis shows absolute angular error in degrees, blue circles show the orientation reports of an observer, the

red line shows the best fitting cumulative gaussian function, green dashed lines show 95% confidence intervals at each optotype size.

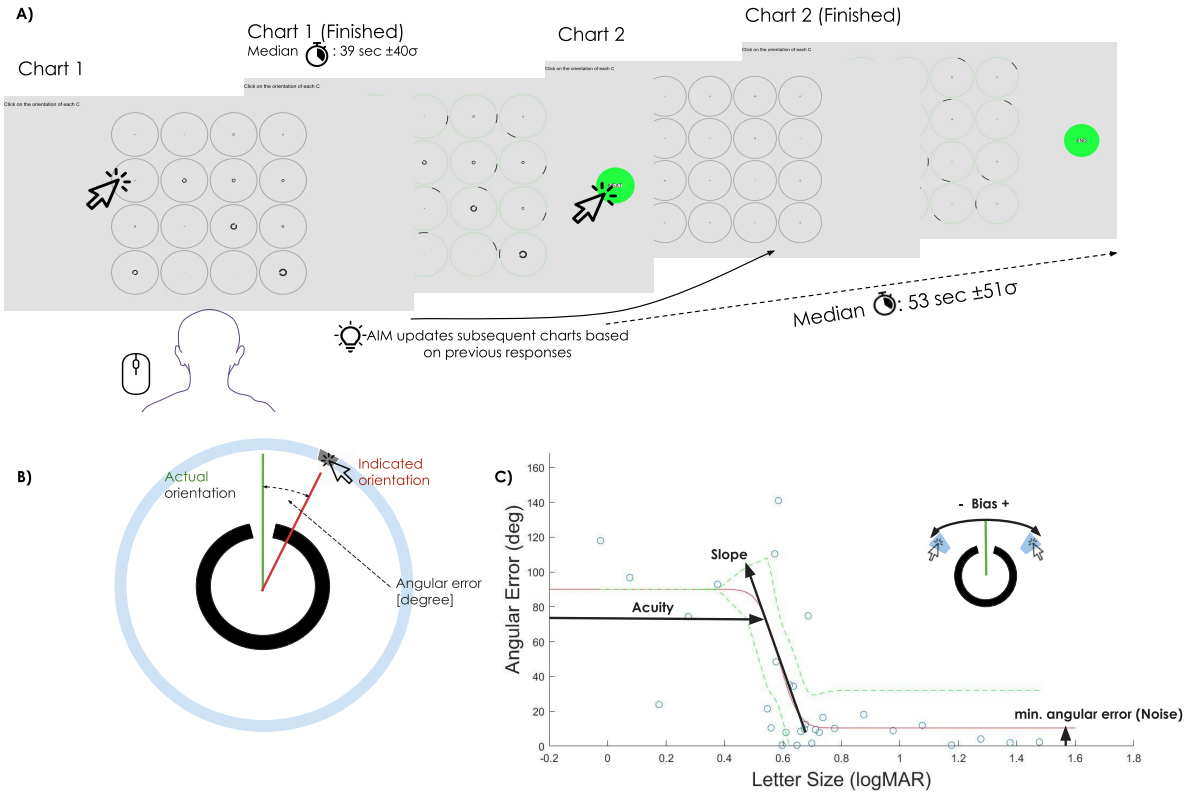

*Figure 1: Schematic of AIM principal. A) Stimuli are presented in series of charts, each containing a 4*4 grid of cells. Each cell consists of a stimulus surrounded by a response ring. Participants reported the orientation of the C optotype in each cell by clicking on the corresponding location on the response ring. Their response is indicated by an arc on the ring, and participants are able to adjust their response by clicking again on the response ring. After a response has been made for all cells, a 'Next' button is presented, which participants can click when they are satisfied with their responses to the chart. The stimulus sizes on the next chart are updated, based on the responses to stimuli in previous charts. B) Scheme of AIM scoring. C) Representative results and fit after two charts. The x axis depicts optotype size in logMAR units, the y axis shows absolute angular error in rotational degrees. Blue circles show the orientation reports of an observer, the red line shows the best fitting cumulative gaussian function, from which acuity was calculated, green dashed lines show 95% confidence intervals at each optotype size. Indicated orientations can be either corresponded with the target orientation (i.e., 0° angular orientation error) or resulting in an orientation error that can either be oriented in clockwise or anticlockwise direction relative to the target orientation.*

Luminance levels of ETDRS, FRACT and AIM

Measured at 1m with a LS 110 photometer, the mean letter and background luminances for ETDRS were 5.2cd/m^2 ±1.3 and 190.6 cd/m^2 ±1.6 and for FrACT were 2.8cd/m^2 ±0.3 and 217.3 cd/m^2 ±5.2, respectively.
The minimum and maximum luminance values of the monitor were 0.3 cd/m^2 and 250 cd/m^2, respectively and gamma was 2.0. The LUT values for the background was adjusted to 248 and 191 to match AIM-VA to FrACT and ETDRS, respectively.

Experiment 1: Comparison between AIM VA, ETDRS, and FrACT

VA for left eye (OS), right eye (OD), and both eyes (OU) was measured at 4m viewing distance for AIM VA, EDTRS, and FrACT in randomized order among participants during constant room lighting conditions. 32 Landolt-C optotypes in 8 possible orientations for a maximum duration of 30sec were shown for FrACT per eye in a randomized order. FrACT size validations where carried out for each test. All other FrACT parameters were its default settings. For the ETDRS chart, each participant was asked to read from the top left letter left to right until the line where fewer than 3 letters were correctly named on a line. On a subset of these data (n=12) we measured the time it took to perform the tasks. Test times for ETDRS and FrACT were measured manually using a stopwatch and for AIM-VA by the Matlab software.

Experiment 2: Test-retest AIM vs ETDRS

In a second experiment, we tested the test-retest repeatability of AIM-VA and compared it to ETDRS test-retest. The dominant eye was tested, and each test was repeated twice on the same day by the same experimenter where the order of tests was randomized. To test whether more adaptive steps for AIM-VA would have an impact on repeatability, we performed a control experiment with a subset of participants who took part in the second experiment and repeated AIM-VA test with four adaptive steps (5 charts).

Data analysis

For ETDRS, VA is defined at the line at which 3/5 (60% correct) letters are correctly identified, which corresponds to an angular error of ±72° for AIM-VA. ETDRS and FrACT have a single parameter estimate (threshold VA), to equate the number of free parameters for all three tests in the current study, we shared the slope and noise level parameters across all data for each subject and allowed threshold to vary in order to have only one free parameter for all tests.

The slope of a psychometric function reveals the change of the detection probability across stimulus levels, and is not typically scored for ETDRS although it can be used to report the logMAR range from guessing to perfect performance (36). AIM provides

a similar range, which we define as the logMAR range between VA and the 99% point on the cumulative Gaussian function, hence 1% above an observers min. angular indication error (Figure 5). We refer to the logMAR distance between VA and the 99% angular error point as *range of stimulus detectability improvement* (ROSDI).

A customized Matlab (MathWorks, R2021a) program was used to analyze the raw data, data fitting, and all statistical tests. For the first experiment, statistical analyses were one- and two-way Analysis-of-Variance (ANOVA) and planned comparisons (*anova1, anovan, multcompare* Matlab functions, respectively). For the analysis of noise, slopes, semi-versus fully constrained, and ROSDI, all available AIM-VA data from Experiment 1 were used. For the second experiment, Bland Altman plots were used in combination with a linear regression model to test for bias.

3. Results

Optometric screening results

First experiment: All participants (age ranging from 18 to 35) had stereo thresholds within a normal range (20” to 100”), showed no suppression on either Random Dot or Worth 4 Dot test. Eye dominance indicated via Miles’ test were 14/15 for OD/OS, respectively. 27 participants were right- and 2 left-handed, 3 participants wore contact lenses, 10 wore glasses, and 16 were without correction.

Second experiment: One participant was excluded due to erroneously duplicated data. Of the remaining 22 participant (age range 18-35 years), 10 wore no optical correction, 5 participants wore contact lenses, and 7 wore glasses. Although no data on subjective, optometrist-performant refraction right before the experiment was available, we assume that participants that wore corrections were best-corrected. For the group without correction worn, the median spherical (sph), negative cylindrical (cyl) values in diopter, axis (x) in degree, spherical equivalent (SE) in diopter, and their respective minimum to maximum values were:

OD-sph: +0.13 D [-0.25, +1.00] / cyl: -0.75 D [-0.75, ±0.00] x 94° [89, 164] | SE: -0.19 D [-0.75, +0.63]
OS-sph: ±0.00 D [-0.25, +1.75] / cyl: -0.38 D [-1.25, ±0.00] x 65° [9, 179] | SE: ±0.00 D [-0.75, +0.63].

A control experiment to test whether more trials would improve the AIM-VA test-retest reliability was performed and included 12 participants. The AIM task was the same as for Experiment two, but with five charts i.e., one initial chart and four adaptive steps.

First experiment- Comparison between AIM, FRACT, and ETDRS

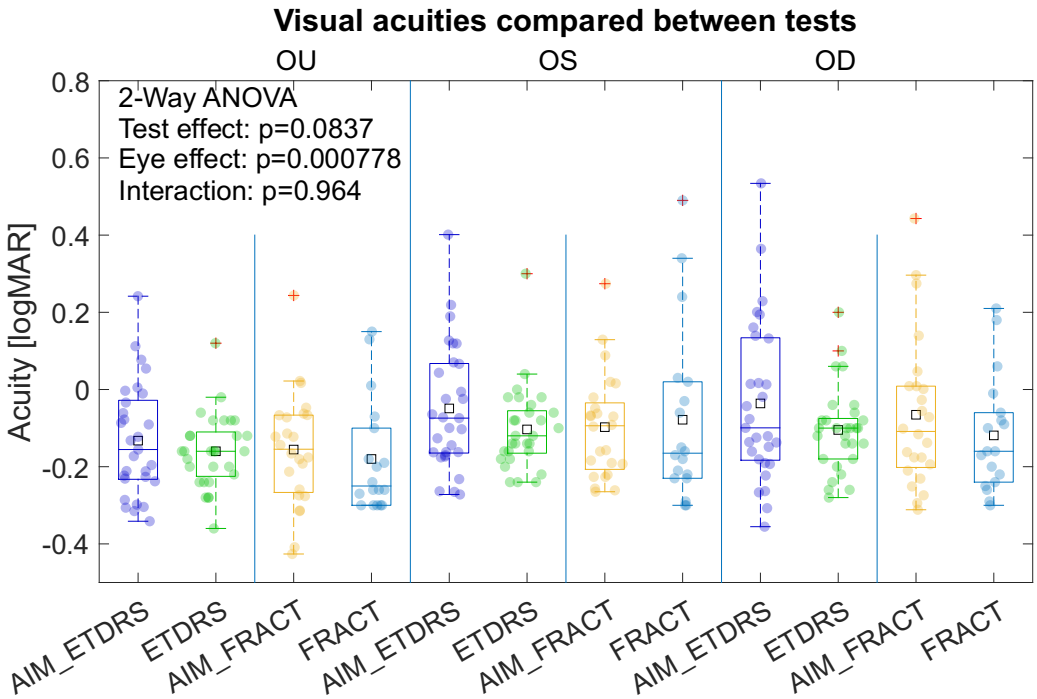

Figure 2: Comparison of visual acuities between AIM luminance-adjusted for ETDRS (violet), ETDRS (green), AIM adjusted for FrACT (beige), and FrACT (blue). Shown are interquartile range (boxes), medians and means (horizonal lines and squares within each box), single outcomes (dots) jitter by using a kernel density estimate, whiskers for extreme values, and red plus for outliers.

A two-way ANOVA comparing test types and eyes (Figure 2) revealed no main effect for test [$F(3,288) = 2.2$, $p>0.05$] and a significant effect of eyes [$F(2,288) = 7.3$, $p<0.001$], which was due to the better (more negative logMAR) acuity for OU relative to each single eye. No interaction was found [$F(6,288) = 1.0$, $p>0.05$]. Planned comparisons between each eye and test showed no significant differences for all comparisons.

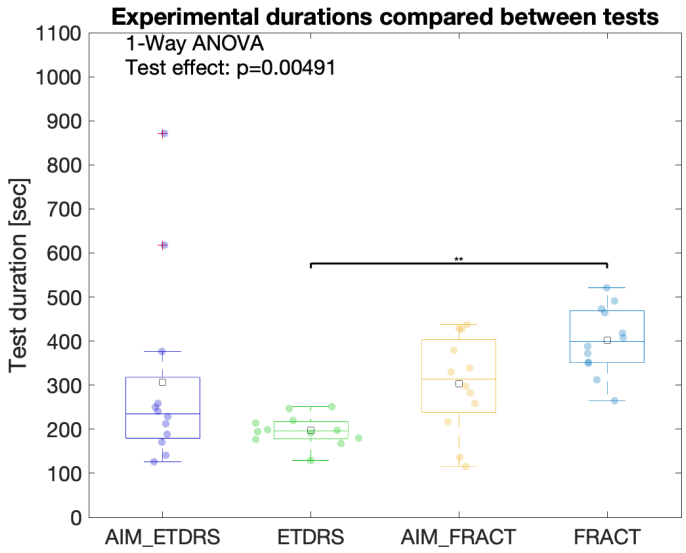

*Figure 3: Comparison of testing times between AIM adjusted for ETDRS comparison, ETDRS, AIM adjusted for FrACT, and FrACT. ** refers to a significant difference (p<0.01) during planned comparison analysis.*

A one-way ANOVA was performed to test for differences in experimental duration between tests (Figure 3) and found a significant effect [$F(3,44) = 4.9$, $p<0.01$]. A planned comparison revealed one significant difference, specifically for a shorter duration of ETDRS compared to FrACT test ($p<0.01$). The median durations measuring all eye conditions were for both AIM tests, ETDRS, and FrACT 259sec ±170 (standard deviation), 197sec ±34, and 399sec ±77, respectively. The median duration for a single eye's AIM tests was 53sec ±51.

Chart durations were also analyzed with a two-way ANOVA across the two AIM-VA test versions and eyes. No significant main effects for AIM test [$F(1,66) = 0.1$, $p>0.05$] and eyes [$F(2,66) = 0.5$, $p>0.05$] nor an interaction [$F(2,66) = 0.1$, $p>0.05$] was found. Median chart duration across tests and eyes was 39sec ±40.

Additional outcomes- Semi-constrained VA, Slopes, Noise, and Range of Stimulus Detectability Improvement

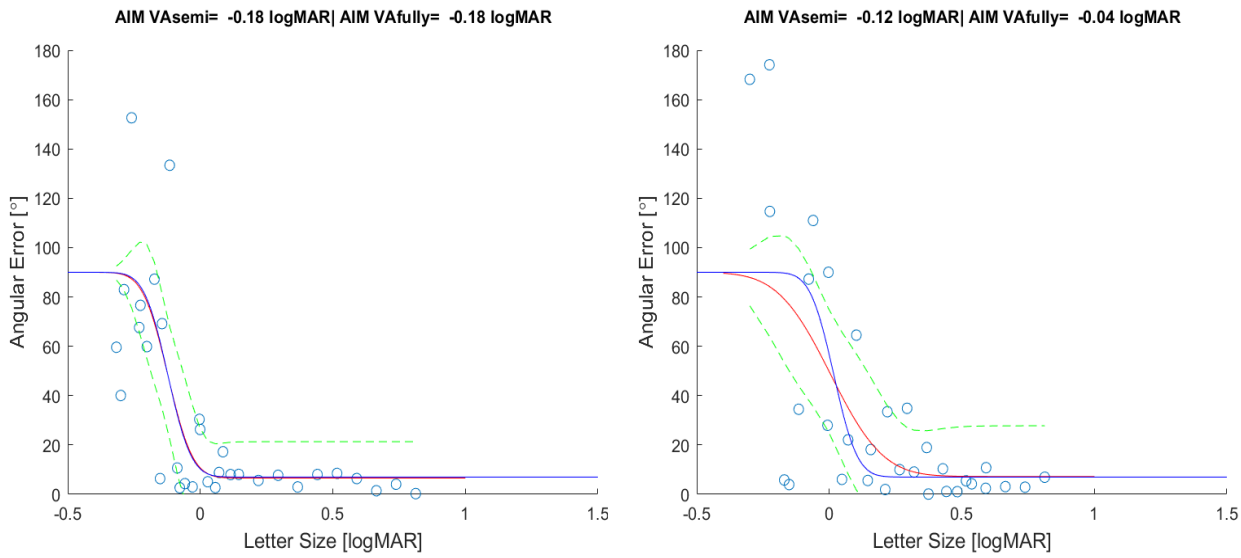

Figure 4: Examples of semi-constrained (red line) with 95% confidence intervals (green dashed lines) and unconstrained (blue line) psychometric function fits for two outcomes within each graph of two different participants (left and right graph).

AIM provides additional outcomes next to VA, namely the slope, noise, and semi-constrained VA's. An example for one observer's AIM VA results with an unconstrained and semi-constrained fit is shown in Figure 4. The VAs in the left example is the same as the unconstrained was identical with that of the semi-constrained fit. The fits on the right however differ such that the VAs differ by 0.08logMAR. We summarized all collected AIM VA in OS, OD, OU from Experiment one for unconstrained and semi-constrained VAs (see supplementary materials). We found no significant main effects and interaction for eyes or AIM tests for both slopes and noise (see supplementary materials). Averaged across participants, AIM types, and eyes, the median slope was 0.07 and the median noise was 7° (see also supplementary materials).

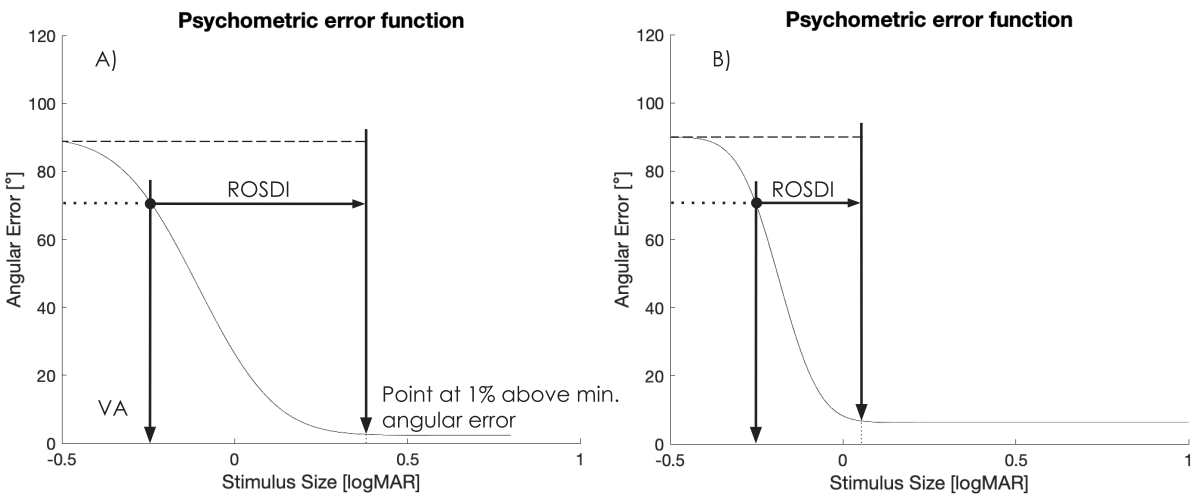

Figure 5: Examples of ROSDI psychometric function fits for two outcomes of two different participants. A) shows ROSDI in an observer with a shallower slope and lower noise, resulting in a larger ROSDI than compared to observer shown in B), whose psychometric function has a steeper slope and higher noise, resulting in a smaller ROSDI.

The ROSDI analysis across all observers (measuring the logMAR distance between threshold (semi-constraint VA) and noise) revealed no significant main effects for AIM test nor eyes, or an interaction of those (see all details in supplementary materials). Averaged across AIM tests and eyes, the median ROSDI was 0.14 logMAR ±0.033.

Second experiment

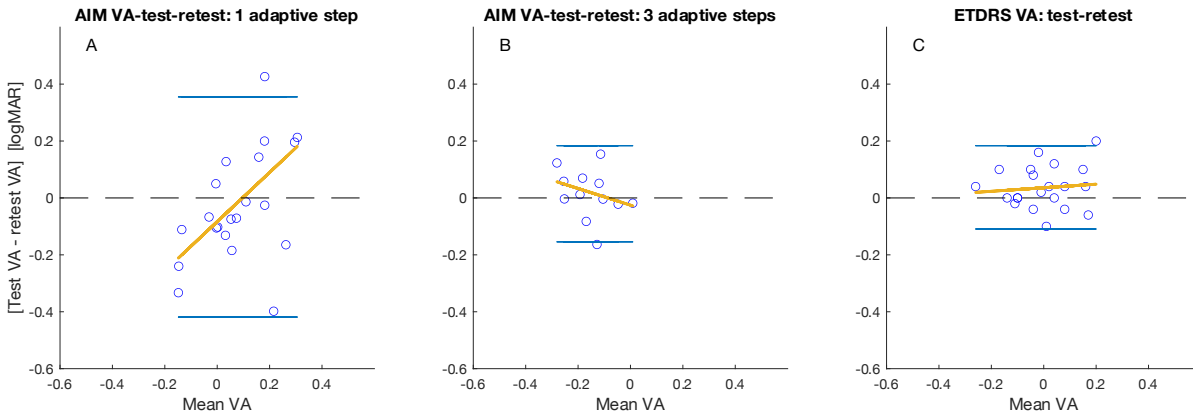

Figure 6: Bland-Altman plots for test-retest reliability comparing AIM VA with an initial chart and one (A) or three adaptive steps (B), and ETDRS (C) with the same single chart repeated once. Each y axis shows the VA test-retest difference in logMAR, each x-axis depicts the mean VA for test and retest VA results. Single circles represent data from individual participants, blue horizontal lines represent 95%

confidence intervals for test-retest VA difference, beige lines represent the linear function fits.

The Bland-Altman plots in Figure 6A show that AIM VA using one adaptive step (2 charts) and ETDRS using the same chart twice (Figure 6C). Test-retest VA differences (y axis) were larger in AIM VA as indicated via upper and lower 95% confidence intervals in green dotted lines but for both tests not significantly biased in either direction ($p > 0.05$). We also fitted a linear regression model (red line) to the data to test for bias across mean VA levels (x axis). The ETDRS showed a significant positive bias($p<0.05$) with an increase with poorer mean VA, which was indicated by a slope that tended to be significantly different from zero line ($p=0.0587$). AIM VA showed no significant test-retest bias, but its slope also trended from to the positive with poorer mean VA ($p=0.07$ for slope). The relative and absolute numbers of test-retest difference for AIM VA were ≤0.1logMAR were 68% (15/22) and ≤0.2logMAR 86% (19/22) and for ETDRS ≤0.1logMAR were 86% (19/22) and ≤0.2logMAR 100% (22/22). The 95% confidence intervals in logMAR for test-retest difference were AIM-VA [-0.36 0.41] and for ETDRS [-0.11 0.18].

As the variance of an AIM fit is dependent upon the amount of data along its slope, its variance should decrease with more sampling along the slope. We ran a control experiment with four adaptive steps (5 charts in total in a subset of participants (n=12) to test this hypothesis. As shown in Figure 6 C, the distance between upper and lower 95% confidence interval in the Bland Altman plots significantly decreased with step sizes [$t(4) = 5.2$, $p<0.01$] and were comparable to that of ETDRS. For all adaptive steps, no bias of mean test-retest VA difference nor slope of the regression function i.e., bias across VAs was found ($p>0.05$). We calculated whether and if so, how much the test-retest difference improves across trials in absolute and relative numbers as well as by determining the confidence intervals (Table 1).

AIM-VA Chart	**AIM-VA Adaptive steps**	**≤ ±0.1 logMAR [% | n]**	**≤ ±0.2 logMAR [% | n]**	**≤ ±0.3 logMAR [% | n]**	**≤ ±0.4 logMAR [% | n]**	**95% Confidence intervals [logMAR]**
1	0	25% | 3/12	58% | 7/12	92% |11/12	92% |11/12	0.44 / -0.37
2	1	42%| 5/12	92% |11/12	100% |12/12	100% |12/12	0.30 / -0.28
3	2	50% |6/12	92% |11/12	100% |12/12	100% |12/12	0.24 / -0.23
4	3	75% |9/12	100% |12/12	100% |12/12	100% |12/12	0.18 / -0.15
5	4	75% |9/12	100% |12/12	100% |12/12	100% |12/12	0.16 / -0.12

Table 1. Comparison of absolute (*n* out of sample size 12) and relative (%) test-retest differences across levels of difference in logMAR as a function of number of adaptive steps and number of charts during the control experiment.

Discussion

AIM results compared to conventional methods

In the first experiment, we compared the self-administered AIM-VA with a single chart ETDRS VA and an 8AFC FrACT VA administered by a researcher. The display used for AIM-VA was equated in terms of background and stimuli luminance to the ETDRS and to the FrACT test. Both AIM-VAs and ETDRS had no time restriction whereas FrACT had a time limit of 30sec for each presented stimulus. There was a main effect of time and specifically that ETDRS was significantly faster than FrACT.

There was no significant difference between the test durations for AIM-VA's and both ETDRS and FrACT tests. Furthermore, we used a semi-constrained fit approach to equate the number of free parameters for AIM FrACT and ETDRS methods. In a control analysis, there was no significant group difference found between AIM unconstrained and a semi-constrained version in which 2 parameters (slope and random error rate) were shared across tests, AIM-VA using a three free-parameter fit generates VA results comparable to those of ETDRS and FrACT.

The positions of AIM-VA's log-spaced stimuli sizes were randomized within each chart whereas ETDRS used log-spaced stimuli sizes that were organized from large to small down the chart. Future studies will examine whether an observer's performance is significantly influenced by the randomization of the positions of letter sizes within each chart.

AIM test-retest results

In experiment two, we compared AIM-VA with ETDRS using a single chart as often deployed in clinical practice. We show that AIM-VA using two charts (i.e., the initial chart and one adaptive step) show greater test-retest variance (Figure 6). This may be due to three reasons, namely a memory bias for a single ETDRS chart, fewer decision options for ETDRS (10 letters) compared with AIM (continuous 360°) options, and AIM-VA's sampling of data along the slope from random report error (90°) to the min. angular error. To address the latter point, we carried out a control experiment whereby we increased the number of adaptive steps to sample more data along the slope area of the psychometric function to improve the fit. We hypothesized that within the time window of the test to retest (within one day), the slope, noise, and threshold of an observer should not significantly change. Indeed, increasing the number of charts improved the test-retest reliability of AIM-VA significantly and was comparable with ETDRS using three or more adaptive steps charts with 4*4 cells (Figure 6).

The test-retest reliability of ETDRS has been reported and 87% of the standard ETDRS chart generated VAs had a ≤0.1logMAR difference and 100% with a ≤0.3logMAR difference (37). Other tests such as FrACT reported comparable 95% confidence intervals of approximately 0.2logMAR with a few outliers (38). AIM-VA showed comparable results when using three or more adaptive steps (table 1). With AIM, the number of cells per chart and the number of adaptive steps can be independently varied. It is unclear how the test-retest repeatability changes by lowering the number of cells-per-charts while increasing the number of adaptive steps. Future research will investigate this question.

Additional AIM outcomes: Semi-constrained VA, Slope, Noise, and Range of Stimulus Detectability Improvement

Clinical standard chart tests deliver threshold values that are known as VAs. Psychophysically, VA is one point on a psychometric function that describes the probability that an observer can detect a signal, in this study gap orientation or letter identity. The stimulus range over which an observer's performance transitions from detecting to not being able to detect the signal is expressed as the slope of the psychometric function, and may vary between observers and may provide important information about the health of the participant, the progression of eye disease or their response to treatment. Practically speaking, the slope describes the rate of transition from seeing to not seeing and can be either very steep or shallow. Chart tests, such as the ETDRS and Snellen charts, do not take this parameter into consideration whereas AIM-VA calculates a personalized slope value and generates a personalized VA using a psychometric function fitting approach (Figure 4).

Previous research demonstrated that the slope of the psychometric function is affected by blur (36), specifically that the slope becomes more shallow with one diopter defocus blur. AIM reports both the slope of the psychometric function and the *range of stimulus detectability improvement*, ROSDI (Figure 5), which had a median value of 0.14logMAR for the observers in the present study. ROSDI is expressed in logMAR and can be converted to all other known VA units. Future research will examine the relationship between this parameter and defocus blur systematically.

Moreover, even though two observers may be able to accurately report a signal in a forced choice paradigm, the quality of their vision may still be distinct between observers or across time. For example, in terms of a conventional letter chart, one observer may correctly name all five letters in all lines before not being able to detect any single letter whereas a second observer may reach the same VA but made mistakes in some of the lines with larger stimuli. Clinically those VAs are identical but the suprathreshold performance differed. Unlike letter charts such as ETDRS, AIM-VA provides an output for that observer-specific behavior, namely the minimum angular error parameter for the most intense (here: largest) stimuli. A number of sources can affect this parameter e.g., hand tremor, eye movements, or neurological

intrinsic variations (spiking variations, neuro-transmitter variation etc.) and structural irregularities in the optics or retina. We termed these sources as *noise* that captures all sources of external and internal noise. Preliminary results for AIM-VA using spherical blur suggests that noise does indeed increase with increased blur (39).

Future directions

As stated in the introduction, VA can be affected due to a variety of reasons, however, the most common reason is refractive error, most likely through defocus or astigmatic blur (1). Our preliminary findings showed that AIM-VA's is able to detect refractive errors (39), but the analysis of the data is extensive and beyond the scope of the current paper, thus will be reported elsewhere. In this paper, AIM has been used to measure VA at 4m distance, which reduces the remaining accommodation required to a quarter diopter. Future studies will examine AIM-VA test-retest reliability for near distances. In the current study, we applied AIM to measure VA, however, due to its generalizability, future studies will exchange stimulus size with other stimulus properties that are processed in distinct neural sites, e.g., color, contrast, stereo disparity, and motion.

5. Conclusions

The Angular Indication Measurement, or AIM, is a response-adaptive, self-administered, and generalizable technique to interrogate visual functions. The current study applied AIM to measure the visual acuities in adults and compared it with standard clinical tests in terms of estimated VA, testing duration, and test-retest reliability. AIM-VA generated VA estimates that were not significantly different from VA estimated with ETDSR and FRACT. Using one adaptive step, AIM showed greater test-retest variance than a single ETDRS chart, which is likely due to memory bias and fewer degrees of variability for the ETDRS (10 letters) chart compared to AIM (continuous 360°). As hypothesized, increasing the number of adaptive steps from 1 to 3 improves the test-retest reliability to the level of a single ETDRS chart. Based on the findings of the current study, it is recommended to use at least 2 adaptive steps to measure distance VA in adults. AIM-VA provides additional information about an observer's principal error level (noise) and slope which and semi-constrained VA outcomes that consider an observers personalized psychometric function. Also, AIM provides a new outcome that describes the range of stimulus detectability improvement, or ROSDI, in clinically useful VA units, e.g., logMAR, and was found to be 0.14logMAR in the current cohort of young, typically sighted participants. Future studies will investigate AIM-VA's ability to detect changes of VA due to refractive errors or neuro-ophthalmic conditions, e.g., amblyopia in both adults and children and apply the AIM principal to other domains of visual perceptual diagnostic e.g., color, contrast, depth, pattern, or motion perception.

Acknowledgment

Supported by NIH R01 EY029713. Portions of this study have been presented at the virtual Vision Science Society Conference 2022 and will be included in a report at the Vision Science Society conference 2023.

Authors contributions

JS and PJB invented and developed AIM. JS developed the overall methodological design of the study. JS, JH, JBS, and NA carried out the optometric screening and collected data. JS wrote the analysis code for the summary statistics. JS wrote the first draft of the manuscript. All authors critically revised the manuscript before submission.

Competing interests

AIM, including AIM Visual Acuity, is disclosed as provisional patented and held by Northeastern University, Boston, USA. JS and PJB are founders and shareholders of PerZeption Inc.

References

1. WHO, *World report on vision* (2019).
2. T. W. Raasch, Spherocylindrical refractive errors and visual acuity. *Optom. Vis. Sci.* **72**, 272–275 (1995).
3. F. G. Holz, *et al.*, Key drivers of visual acuity gains in neovascular age-related macular degeneration in real life: Findings from the AURA study. *Br. J. Ophthalmol.* **100**, 1623–1628 (2016).
4. J. H. Shandiz, *et al.*, Effect of cataract type and severity on visual acuity and contrast sensitivity. *J. Ophthalmic Vis. Res.* **6**, 26–31 (2011).
5. W. Helmut, S. Martin, The diagnosis and treatment of optic neuritis. *Dtsch. Arztebl. Int.* **112**, 616–626 (2015).
6. D. M. Levi, Rethinking amblyopia 2020. *Vision Res.* **176**, 118–129 (2020).
7. P. T. V. M. de Jong, A history of visual acuity testing and optotypes. *Eye*, 1–12 (2022).
8. H. Küchler, “Schriftnummerprobe für Gesichtsleidende” (1843).
9. M. D. Sheridan, H. Consul-, Sheridan-Gardiner Test for Visual Acuity. *Br. Med. J.* **2**, 108–109 (1970).
10. L. M. Aiello, F. L. Ferris, Photocoagulation for Diabetic Macular Edema. *Arch. Ophthalmol.* **103**, 1796–1806 (1985).
11. F. L. Ferris, A. Kassoff, G. H. Bresnick, I. Bailey, Visual acuity charts for clinical research. *Am. Orthopt. J.* **Vol. 36**, 14–18 (1986).
12. H. Snellen, *Letterproeven, totbepaling der gezigtsscherpte*, J. G. Utrech, Ed.

(P. W. van de Weijer, 1862).

13. H. Snellen, On the methods of determining the acuity of vision. *Syst. Dis. eye* **2**, 93–98 (1897).

14. Landolt E, Nouveaux opto-types pour la détermination de l'acuité visuelle. *Arch d Ophtal* **19**, 465–71 (1899).

15. L. HYVÄRINEN, R. NÄSÄNEN, P. LAURINEN, New Visual Acuity Test for Pre-School Children. *Acta Ophthalmol.* **58**, 507–511 (1980).

16. L. M. Hamm, J. P. Yeoman, N. Anstice, S. C. Dakin, The Auckland Optotypes: An open-access pictogram set for measuring recognition acuity. *J. Vis.* **18**, 1–15 (2018).

17. H. Strasburger, J. Huber, D. Rose, Ewald Hering's (1899) On the Limits of Visual Acuity: A Translation and Commentary: With a Supplement on Alfred Volkmann's (1863) Physiological Investigations in the Field of Optics. *Iperception.* **9** (2018).

18. J. Atkinson, O. Braddick, E. Pimm-smith, "Preferential looking" for monocular and binocular acuity testing of infants. *Br. J. Ophthalmol.* **66**, 264–268 (1982).

19. I. L. Bailey, J. E. Lovie, New Design Principles for Visual Acuity Letter Charts. *Am. J. Optom. Physiol. Opt.* **53**, 740–745 (1976).

20. G. T. Fechner, *Elemente der Psychophysik* (Breitkopf und Härtel, 1860).

21. C. W. Mcmonnies, C. A. Hazel, D. B. Elliott, The Dependency of LogMAR Visual Acuity Measurements on Chart Design and Scoring Rule. **80**, 486–489 (2003).

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

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

24. M. Bach, The Freiburg Visaul Acuity Test- Automatic Measurment of Visual Acuity. *Optom. Vis. Sci.* **73**, 49–53 (1996).

25. C. Piech, A. Malik, L. M. Scott, R. T. Chang, C. Lin, The stanford acuity test: A precise vision test using bayesian techniques and a discovery in human visual response. *AAAI 2020 - 34th AAAI Conf. Artif. Intell.*, 471–479 (2020).

26. E. R. F. W. Crossman, P. J. Goodeve, A computer-based automatic method for determing visual acuity. *Optom. Vis. Sci.*, 344–355 (1970).

27. X. Han, *et al.*, Development and validation of a smartphone-based visual acuity test (Vision at home). *Transl. Vis. Sci. Technol.* **8** (2019).

28. A. Samanta, S. Mauntana, Z. Barsi, B. Yarlagadda, P. C. Nelson, Is your vision blurry? A systematic review of home-based visual acuity for telemedicine. *J. Telemed. Telecare* (2020) https:/doi.org/10.1177/1357633X20970398.

29. W. K. Yeung, *et al.*, eHealth tools for the self-testing of visual acuity: a scoping review. *npj Digit. Med.* **2** (2019).

30. S. Akbar, E. Coiera,, F. Magrabi, Safety concerns with consumer-facing mobile health applications and their consequences: A scoping review. *J. Am. Med. Informatics Assoc.* **27**, 330–340 (2020).

31. Consilium Ophthalmologicum Universale, VISUAL ACUITY MEASUREMENT STANDARD.

32. D. G. Pelli, The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spat. Vis.* **10**, 437–442 (1997).

33. D. H. Brainard, The Psychophysics Toolbox. *Spat. Vis.* **10**, 433–436 (1997).

34. J. Skerswetat, P. J. Bex, AIM + (Angular Indication Measurement Plus) enables rapid , self-administered assessment of visual perception dependency across multiple stimulus dimensions in *European Conference for Visual Perception*, (2022).

35. J. He, J. Skerswetat, P. J. Bex, Novel color vision assessment tools: AIM Color Detection and Discrimination in *Association for Research in Vision Science and Opthalmology and Opthalmology*, (2023).

36. A. Carkeet, L. Lee, J. R. Kerr, M. M. Keung, The slope of the psychometric function for Bailey-Lovie letter charts: Defocus effects and impilications for modeling letter-by-letter scores. *Optom. Vis. Sci.* **78**, 113–121 (2001).

37. R. W. Beck, *et al.*, A computerized method of visual acuity testing: Adaptation of the early treatment of diabetic retinopathy study testing protocol. *Am. J. Ophthalmol.* **135**, 194–205 (2003).

38. M. Bach, The Freiburg Visual Acuity Test-Variability unchanged by post-hoc re-analysis. *Graefe's Arch. Clin. Exp. Ophthalmol.* **245**, 965–971 (2006).

39. M. Freeman, J. Skerswetat, P. J. Bex, Measuring the Effect of Negative and Positive Defocus on Visual Acuity Using AIM (Angular Indication Measurement) Acuity Key Messages in *Virtual Vision Science Society Conference*, (2022), p. Poster.