

Automated Measurement of Resolution Acuity in Infants Using Remote Eye-Tracking

Pete R. Jones,¹ Sarah Kalwarowsky,¹ Janette Atkinson,^{2,3} Oliver J. Braddick,³ and Marko Nardini^{1,4}

¹Institute of Ophthalmology, University College London (UCL), United Kingdom

²Department of Developmental Science, University College London (UCL), United Kingdom

³Department of Experimental Psychology, University of Oxford, United Kingdom

⁴Department of Psychology, Durham University, United Kingdom

Correspondence: Pete R. Jones; Institute of Ophthalmology, University College London, 11-43 Bath Street, London EC1V 9EL, UK; p.r.jones@ucl.ac.uk

Submitted: June 25, 2014

Accepted: October 10, 2014

Citation: Jones PR, Kalwarowsky S, Atkinson J, Braddick OJ, Nardini M. Automated measurement of resolution acuity in infants using remote eye-tracking. *Invest Ophthalmol Vis Sci.* 2014;55:8102-8110. DOI:10.1167/iov.14-15108

PURPOSE. To validate a novel, automated test of infant resolution acuity based on remote eye-tracking.

METHODS. Infants aged 2 to 12 months were tested binocularly using a new adaptive computerized test of infant vision using eye tracking (ACTIVE), and Keeler infant acuity cards (KIAC). The ACTIVE test ran automatically, using remote eye-tracking to assess whether the infant fixated a black-and-white grating of variable spatial frequency. Test-retest reliability was assessed by performing each test twice. Accuracy was assessed by comparing acuity measures across tests and with established age-norms, and by comparing low-contrast acuity estimates in adults with data reported previously.

RESULTS. All infants completed the ACTIVE test at least once. Median test duration was 101 seconds. Measured visual acuity increased with age ($P < 0.001$), and 90% of mean acuity estimates were within previously published 90% tolerance limits (based on acuity-card age norms). Acuity estimates were also correlated, within-subjects, with results from the KIAC ($P = 0.004$). In terms of reliability, 86% of acuity estimates deviated by ≤ 1 octave, with no significant difference in test-retest reliability between the ACTIVE and KIAC procedures ($P = 0.461$). In adults, acuity estimates from the ACTIVE test did not differ significantly from values reported by previous authors ($P > 0.183$).

CONCLUSIONS. An adaptive computerized test of infant vision using eye-tracking provides a rapid, automated measure of resolution acuity in preverbal infants. The ACTIVE performed comparably to the current clinical gold standard (acuity cards) in terms of testability, reliability, and accuracy, and its principles can be extended to measure other visual functions.

Keywords: visual acuity, infant vision, eye-tracking

Quick and accurate behavioral measures of resolution acuity are vital for assessing vision, both in clinical practice and in research. These measures can be used to detect pathologies,¹⁻³ predict visual outcomes,^{4,5} and assist in the planning and assessment of treatments.⁶ The ability to measure acuity during infancy is particularly important, since this is when the visual system is developing most rapidly,⁷ and interventions may be most effective.^{8,9}

In infants, the current gold-standard test of functional acuity is the preferential-looking acuity card procedure.¹⁰ The operator presents the infant with a sequence of cards, each containing a black-and-white grating on either the left or right side. Gratings vary in spatial frequency, and are presented against a gray background of matched mean luminance. Given their preference for pattern over uniformity,¹¹ infants will tend to fixate the grating pattern if they can resolve it. A trained operator judges whether the infant fixates the grating, and determines the highest spatial frequency that they fixate reliably.

The acuity card procedure has changed little since its introduction, 35 years ago.¹²⁻¹⁵ This lack of development partly reflects its effectiveness. Acuity cards yield results in $\sim 95\%$ of

healthy infants,¹⁶⁻¹⁸ and within 5 minutes can give estimates of acuity that, in $\sim 90\%$ of cases,¹⁶ are reliable to within 1 octave^{12,18-20} (i.e., a doubling or halving of spatial frequency).

However, acuity cards do have limitations. A substantial practical drawback is that they require an expert operator. Thus, despite their apparent simplicity, effective use of the cards demands "a practiced clinician"²¹ with "considerable experience"⁶ and "considerable judgment."¹⁶

In clinical environments, the operator often has little knowledge of, or control over, key test parameters. In particular, the luminance of the test card in the lighting conditions of the room, its presentation distance, or precisely where in the visual field the grating stimulus is presented. For example, while presentation distance should be constant, it is common for infant, parent, or operator to move during testing. Such movements are known to occur even when testing acuity in adults,²² and given that acuity cards are presented in the near-field (i.e., at 38-84 cm), a movement of 10 cm could cause acuity measurements to vary by 25%. Similarly, since the infant's initial fixation position cannot be controlled, the location of the stimulus within the visual field is liable to vary across trials, again causing expected acuity to fluctuate.^{23,24}

Such variations in luminance, distance, or position, are potential sources of measurement error, and may also bias results systematically. For example, differences in luminance have been hypothesized to explain differences in acuity of up to 1 octave between laboratories.^{25,26} Other factors, such as the level of experience of the operator (Brown A, unpublished observations, 2014), may further affect the acuity estimates, and are difficult to quantify or report.

The optimal stimulus for acuity testing is a sine-wave grating, contrast modulated smoothly at the edges. This is difficult to achieve with printed cards, which are therefore liable to exhibit edge effects. Further confounds may be introduced over time as the cards become scratched or faded. Such artifacts are likely to have negligible impact in very young infants, but are clearly visible by adulthood, where they can be used as cues to perform the task. Their effect on older infants and children is unknown.

Finally, acuity cards are inflexible. In some cases, it may be advantageous to vary the range or distribution of test stimuli (e.g., in order to track small changes over time). Similarly, in some circumstances it may be beneficial to vary the contrast,^{27,28} hue,^{29,30} or spatial location^{31,32} of the stimuli in order to more fully characterize the infant's visual system. These kinds of modifications cannot be implemented practically using printed cards.

In the present work, we addressed these challenges by developing a novel, computer-based system in which stimuli are displayed on an LCD screen, while a remote eye-tracker precisely tracks infants' looking responses. The result is the ACTIVE test, suitable for use with nonverbal observers. The protocol described here measures resolution acuity, but the same principles—combining remote eye-tracking with automated algorithms—can be extended to measure other aspects of visual function, such as contrast sensitivity, chromatic discrimination, field testing, and spatiotemporal sensitivity.

METHODS

Participants

Observers were 30 infants (16 female), aged 2.6 to 12.7 months, with no known visual problems or medical conditions. Ages were corrected for gestational age at birth, and were distributed approximately uniformly (the age distribution can be seen in the “Results” section, where acuity is plotted against age). No infant was born more than 13 days premature, and mean age at birth was 2.7 days postterm. Birth weight was not recorded, but all children had normal birth histories, and none required special neonatal care. Infants were recruited via advertisements in maternity wards and the local area.

Eight adults (five female) were also tested, using a low (10%) contrast version of the same test. Adults were aged 21.6 to 39.9 years, had normal or corrected-to-normal vision, were recruited through the UCL psychology subject pool, and received £7 compensation for their time.

Written informed consent was obtained from parents (infants) or participants (adults) before testing. The research was carried out in accordance with the tenets of the Declaration of Helsinki, and was approved by the local National Health Service England ethics committee (infants), and the UCL Ethics Committee (adults).

Apparatus

The equipment for the ACTIVE procedure is shown in Figure 1A. It consisted primarily of a remote eye tracker (Tobii TX120; Tobii Technology AB, Danderyd, Sweden) operating at 60 Hz,

and a 30-inch liquid-crystal display (LCD) monitor (Samsung SyncMaster 305T, 64 × 40 cm screen, 2560 × 1600 pixels, 0.25-mm dot pitch, 60 Hz; Samsung Electronics Co., Ltd., Seoul, South Korea) interfaced using a graphics card (Nvidia GeForce GTX 650Ti; Nvidia Corp., Santa Clara, CA, USA). Hardware were controlled with custom code written in a numerical computing environment (MATLAB; R2012b, MathWorks, Natick, MA, USA), using commercial software (Psychophysics Toolbox, v3^{33,34}; MathWorks, and Tobii Analytics SDK, v3; Tobii Technology AB). The eye tracker was calibrated independently during each test, using a novel procedure described in the Supplementary Material. A standard set of Keeler infant acuity cards (KIAC; Keeler Ltd., Windsor, UK) was also used to measure acuity, as detailed below.

Stimuli

The stimuli for the ACTIVE test were horizontal Gabor patches: stationary sine-wave gratings, modulated by a Gaussian window. Use of a sine-wave grating ensured that the spatial frequency of the stimulus was precisely controlled, and the Gaussian modulation ensured that the edge of the stimulus could not be used as a detection cue. The standard deviation of the Gaussian hull was 1.47°. Thus, 95% of the signal energy fell within a circle, 5.9° in diameter.

Gratings were presented at 99.6% Michelson contrast for infants (the maximum, given the hardware), and at 10% Michelson contrast for adults. At 99.6% contrast, the space-averaged luminance of the target was 87.43 to 101.5 cd/m² (depending on the position on the screen; see Supplementary Material). Gratings were presented against an isoluminant gray background (adjusting for target location, so that the mean luminance of the target always matched the local mean luminance of the background). Any gaze coordinates that fell within a 7.2° × 7.2° square (~10 cm²) centered on the stimulus were considered on target (see Classifying Hits/Misses). This area represents three times standard deviation of the Gabor, plus a uniform border of 1.4°, and was found during piloting to provide a reasonable compromise between hits and false alarms.

Note that in neither the ACTIVE nor KIAC procedures is the grating stimulus presented foveally. Both tests are therefore best thought of as measuring paracentral acuity, and provide a lower bound on “best” acuity (note that the exact development of acuity across the visual field is unknown, but by 2 months, resolution acuity appears greatest at the fovea³⁵). The exact eccentricity of the grating in the two tests cannot be compared straightforwardly, but was generally more central in the ACTIVE test (see Supplementary Material).

Acuity Cards

Keeler infant acuity cards were also used as an independent, within-subjects measure of acuity. Keeler cards are the most common measure of infant acuity in UK clinics, and have been shown to yield results comparable with the original Teller acuity cards³⁶ (for which extensive normalization data are available, but which can no longer be purchased, having been replaced by Teller II cards³⁷).

The Keeler infant acuity cards consists of seven cards (plus one blank), each containing a square wave grating in one of two locations. At the test distance of 38 cm, the spatial frequencies of the gratings were: 0.18, 0.36, 0.72, 1.4, 2.9, 6.5, or 12.5 cycles per degree (cpd). The test room was well lit, and the mean level of illumination on the cards was approximately 40 cd/m² (measured by a CS-100A Chroma Meter; Minolta Camera Co., Osaka, Japan). This was well in excess of the minimum level recommended by the manufacturer²⁶ (10 cd/

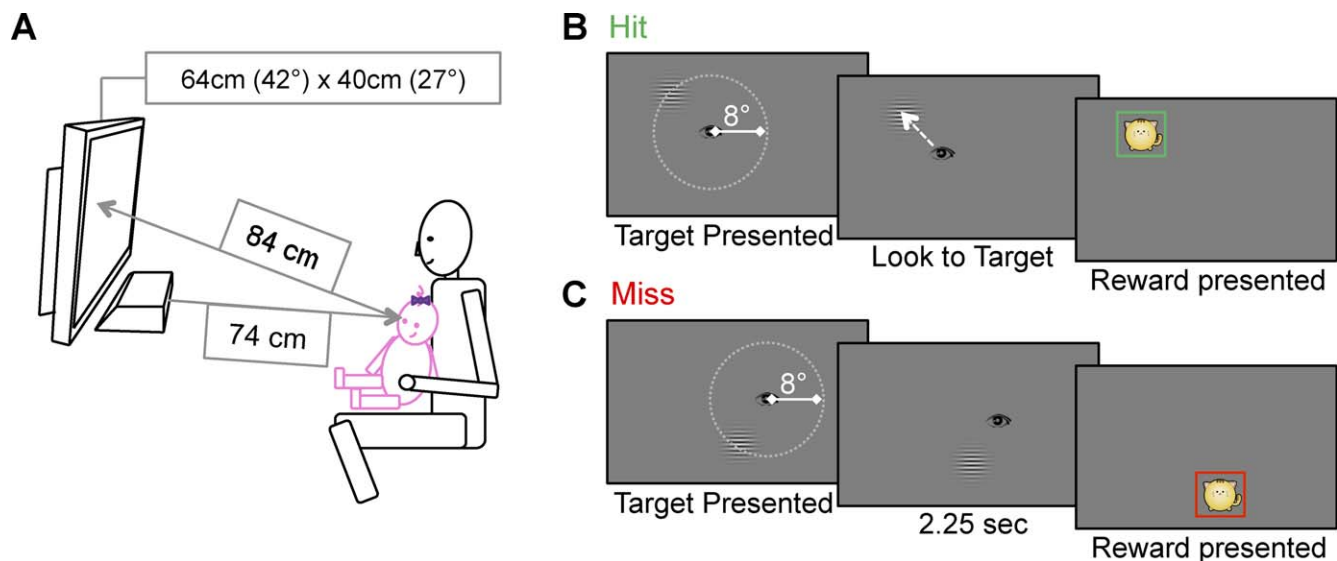


FIGURE 1. ACTIVE apparatus and procedure. (A) The infant was seated on a parent's lap, and viewed stimuli binocularly at a distance of 84 cm. An eye-tracker (Tobii Technology) was mounted below the LCD screen, and recorded the infant's eye movements. (B) On each trial, the screen was initially blank. A single Gabor grating was presented, the center of which was located at a random location, 8° of visual angle from the infant's initial point of fixation. Infants had 2.25 seconds to look to the target (hit), otherwise, the trial was scored as a miss. (C) A visual "reward" was presented at the stimulus location at the end of every trial (independent of performance).

m^2), and within the range of values used in previous similar studies^{10,20} (16–56 cd/m^2).

The operator (PJ or SK) was blind to the target location, but because of the adaptive procedure, was not blind to the spatial frequency. The parent holding the infant was not blind to the target location, in either the ACTIVE or KIAC procedures. However, there was no indication that this affected infants' performance. This was assessed informally by asking parents to close their eyes periodically during testing. Any parental cues would be particularly unlikely to affect scoring in the ACTIVE, since a "hit" required the infant to fixate a precise and variable screen location (e.g., as opposed to a simple left/right eye movement). Each operator had experience running the KIAC test in approximately 30 infants prior to testing, and the data showed no evidence of improving acuity over the course of collection—suggesting that the operators' skill at using the KIAC procedure had reached asymptote.

Because the results of the KIAC procedure appeared to be of questionable validity (see "Discussion" section) post hoc, the present data were also compared with previous, normative datasets collected using Teller acuity cards.

Procedure

During testing, the infant was seated on a parent's lap and viewed stimuli binocularly at a distance of 84 cm (ACTIVE), or 38 cm (KIAC). Binocular viewing was used in order to avoid dropout rates being confounded by the process of patching. However, monocular testing is also possible using both techniques (see "Discussion" section). Test distance was established initially by aligning the infant with a set of premeasured marks, and was enforced in the ACTIVE procedure by pausing the test if the infant moved by 10 cm or more in any direction (as estimated by the eye-tracker). The farther test distance in the ACTIVE procedure was necessitated by the eye-tracker's depth of focus. It is unlikely to have affected the results, since 3- to 12-month-old infants are capable of accommodating at this distance,³⁸ and systematically varying viewing distance from 30 to 150 cm has been found not to affect acuity estimates in 1- to 2-month-olds.³⁹

However, nearer viewing may be necessary when testing infants in the first 2 months of life,³⁸ to ensure attention and accurate accommodation.

Infants were tested twice using each measure, in an alternating ABAB sequence. First test performed (ACTIVE or KIAC) was counterbalanced across subjects. Breaks were taken between tests as required. Each test lasted approximately 1.5 minutes (see "Results" section), and sessions typically lasted 30 minutes in total.

We performed the KIAC procedure in accordance with the manufacturer's instructions. Starting at 0.36 cpd, the operator presented each card by holding it in front of his or her face.⁴⁰ Spatial frequency increased in 1-octave steps until the operator was unable to determine the grating location based on the infant's behavior. The previous spatial frequency was then repeated for confirmation, and this was taken as the acuity threshold. Further trials were employed as required in the event of inconsistent responses.

The ACTIVE procedure was conceptually similar, but ran automatically, without any intervention from the experimenters. On each trial, the screen was initially blank (uniform gray). A grating of set spatial frequency was then presented against this isoluminant background, and an automated algorithm determined whether the infant fixated it (see Classifying Hits/Misses). The infant's gaze was not directed to the center of the screen prior to each trial. Instead, the center of the stimulus was positioned at a random location, 8° from the infant's point of fixation (measured at trial-onset), with the constraint that the stimulus's center had to lie $>3.6^\circ$ from any screen edge. This meant that the range of possible stimulus locations was reduced if the infant began the trial fixating eccentrically. On trials where no eye tracking data were available at trial onset (e.g., if the infant had turned away or closed his or her eyes), then the target was placed in the center of the screen. In practice, this happened in only 1% of trials, since trials were automatically paused if the infant turned away or moved outside the trackable area.

Across trials, a weighted "up-2 down-1" staircase⁴¹ was used to adapt the spatial frequency of the grating. Starting at 0.88 cpd, spatial frequency increased by 1 octave after a

successful look (hit, defined below), and decreased by 0.5 octaves otherwise (miss). This strategy targets the 33.3% correct point on the psychometric function, and was found to be the most robust during piloting (note that the false-alarm rate on this task was below 10%, as reported in the “Results” section). The lowest permissible spatial frequency was 0.61 cpd. The highest spatial frequency was 15.3 cpd, after which point the trial proceeded as normal, but no grating was actually displayed (blank trial). The adaptive staircase continued until at least 15 trials (mean = 18.0) and six reversals (mean = 7.4) had occurred. For comparison with the acuity cards, threshold acuity was defined as the highest spatial frequency that the infant was scored as having looked at. However, qualitatively similar results could be obtained by geometrically averaging the last n reversals.

Cartoons were presented after every three trials to maintain interest. The cartoon played for a minimum of 8 seconds and terminated only if/when the infant’s gaze location was registered by the eye tracker (which required them to be looking toward the screen). A “reward” was presented after each trial, irrespective of performance, in the form of a colored animal graphic that appeared at the location of the target, together with an associated sound. A green (“hit”) or red (“miss”) box around the reward provided feedback for the experimenter on how the trial had been scored (see Fig. 1).

Classifying Hits/Misses

The infant was judged to have looked at the stimulus (a “Hit”) if 10 gaze samples (167 ms) fell in a $7.2^\circ \times 7.2^\circ$ square, centered on the target. This corresponds to $\sim 3\%$ of the screen area and as such, the task might be considered a “33 alternative forced-choice” (33AFC). However, this characterization is imprecise; since no decision was forced, an observer could fixate multiple locations, and the spatial distribution of looking when no stimulus was not uniform. The chance (false positive) rate on this task was therefore determined empirically, by presenting a single “invisible target” trial at the end of each test (see “Results” section). These catch trials proceeded exactly as normal, except that no target was actually displayed. Each catch trial was preceded by a suprathreshold (0.61 cpd) trial, intended to maximize attentiveness.

If no hit was detected within 135 gaze samples (2.25 seconds), then the trial was scored as a miss. Note that trials could last longer than 2.25 seconds if the eye tracker failed to detect the infant’s gaze for a period of time (e.g., if infants turned away or covered their eyes). For example, 125 trials (9%) lasted more than 2.5 seconds. Trial durations were found to be appropriate for the age range tested, but longer durations may be required to avoid false negatives when testing younger infants, or children with neurological pathologies.⁴²

Adults

Adults were tested using the same ACTIVE procedure as the infants. However, stimulus contrast was attenuated to 10%: a level expected to avoid ceiling effects in adults, given our range of spatial frequencies.^{23,24} To maximize comparability with the infants, adult observers were instructed only to “sit in front of the screen and relax.”

Analysis

For statistical analyses, acuity thresholds were \log_2 transformed, and test-retest differences in acuity were expressed in octaves rather than raw cpd.⁴³ All parametric tests were two-tailed.

RESULTS

Testability and Test Duration

Twenty-nine of thirty infants (97%) completed two test runs of both the ACTIVE and KIAC procedures. One additional infant (aged 4.1 months) completed only one run of each before becoming uncooperative.

There was no significant difference in test duration (Fig. 2) between the ACTIVE (median = 101 seconds) and KIAC (median = 108 seconds) procedures (paired t -test of log-transformed data; $t_{175} = -1.57$, $P = 0.118$, ns). Note, however, that the average trial duration was substantially faster with the ACTIVE test ($\mu_{\text{ACTIVE}} = 1.92$ seconds; $\mu_{\text{KIAC}} \approx 10$ seconds), which allowed over twice as many trials to be collected within a similar timeframe.

In the ACTIVE test, there was no consistent relationship between test duration and number of trials completed ($r_{57} = 0.21$, $P = 0.116$, ns). However, some tests did take substantially longer than average (max = 208 seconds; see Fig. 2). This was primarily because of brief pauses within some tests (e.g., these occurred automatically if the infant turned away). There was also no relationship between test duration and estimated acuity ($r_{57} = -0.16$, $P = 0.220$, ns).

There was no effect of age on test time, either with the ACTIVE test ($r^2 = 0.01$; $F_{(1,28)} = 0.37$, $P = 0.550$, ns) or the KIAC test ($r^2 = 0.03$; $F_{(1,28)} = 0.80$, $P = 0.377$, ns). No indication of fatigue was apparent in the timing data. With the ACTIVE procedure, overall test times ($t_{28} = 2.49$, $P = 0.019$) and mean trial durations ($t_{28} = 2.20$, $P = 0.036$) both decreased significantly between the first and second run. With the KIAC procedure, only overall test time was recorded, and this did not differ significantly across the two runs ($t_{28} = -0.38$, $P = 0.709$).

Measured Acuity

Visual acuity in the ACTIVE test increased with age ($r^2 = 0.43$; $F_{(1,57)} = 43.04$, $P < 0.001$), and there was good agreement with normative data acquired previously using Teller acuity cards (Fig. 3). For example, 46 (78%) of the individual estimates, and 27 (90%) of the within-subject means lay within the 90% tolerance limits for binocular acuity reported by Salomão and Ventura.¹⁸

The rate at which acuity develops during the first year can be measured by the slope of a linear regression (acuity against age). For example, in Salomão and Ventura’s¹⁸ data, acuity improved at a rate of 1.0 cpd/mo. Analysis of our mean data yielded a similar developmental improvement of 0.98 cpd/mo. This was not significantly different from 1.0 ($t_{28} = -0.38$, $P = 0.710$, ns), though it was slightly greater than the rates of improvements observed by Teller et al.¹² under binocular viewing conditions (0.63 cpd/mo; $t_{28} = 2.19$, $P = 0.037$), or by Mayer et al.²⁰ under monocular conditions (0.58 cpd/mo; $t_{28} = 2.53$, $P = 0.017$).

As has been reported previously,³ there was no difference in the acuity of male and female infants (partial regression, accounting for age; $t_{27} = 0.39$, $P = 0.697$, ns).

The estimates of the KIAC test also increased with age ($r^2 = 0.08$; $F_{(1,57)} = 5.12$, $P = 0.028$). However, the rate of improvement (0.24 cpd/mo) was significantly less than was observed here using eye-tracking ($t_{28} = -3.53$, $P < 0.001$), or has been reported previously (all $P < 0.01$).

Accordingly, while the mean acuity thresholds of the KIAC and ACTIVE tests were robustly correlated ($r_{28} = 0.51$, $P = 0.004$), the KIAC gave lower estimates in 87% of individuals (Fig. 4B). As detailed in the discussion, this mismatch can potentially be attributed to the test frequencies used in the KIAC, the values of which are depicted by black arrows in Figure 4A.

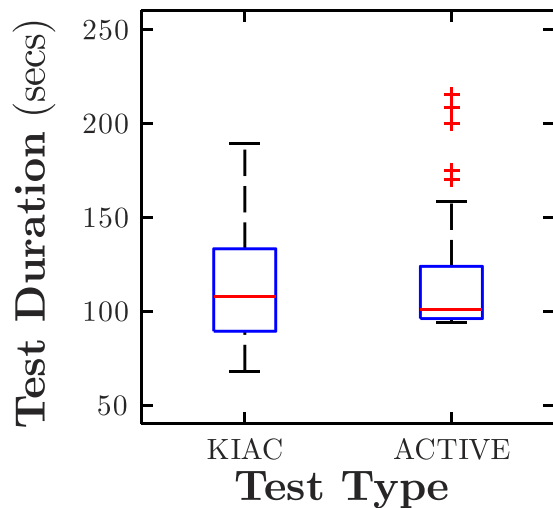


FIGURE 2. Box-plot of test durations ($n = 59$) using the KIAC and ACTIVE procedures. Each test required an additional period of approximately 30 seconds in which to position the infant appropriately.

Test-Retest Reliability

Across two runs, ACTIVE acuity estimates varied by a mean of 0.04 octaves (nonsignificant increment: $t_{28} = 0.19$, $P = 0.851$, ns), with 86% of points varying by 1 octave or less (Fig. 5). The mean test-retest difference of the KIAC procedure was -0.12 octaves (nonsignificant decrement: $t_{28} = -1.01$, $P = 0.319$, ns), with 96% of points varying by 1 octave or less. Although the KIAC test exhibited less within-subject variability, this difference was not significant (paired t -test of Δ_{acuity} across runs; $t_{28} = 0.75$, $P = 0.461$, ns), indicating that neither test was substantially more reliable.

Effects of Target Location

The proportion of correct responses was not affected by where on the screen the grating was presented (Fig. 6A), or by

the grating's relative location in the visual field (Fig. 6B). This was assessed formally by computing, for each participant, proportion correct in each of four location quadrants (Figs. 6A, 6B, dashed black lines). One-way repeated-measures ANOVAs were then performed, with grating location (quadrants 1–4) as a factor. There was no significant effect of either absolute (Fig. 6A, $F_{(3, 87)} = 0.20$, $P = 0.897$, ns) or relative (Fig. 6B, $F_{(3, 87)} = 1.36$, $P = 0.261$, ns) grating location on proportion correct. This demonstrates that neither the infant, nor the eye-tracker, varied in sensitivity across test locations.

Guessing Rate

Analysis of “invisible target” (catch) trials indicated that the target region was fixated by chance on 7% of trials. This is roughly equivalent to the expected guessing (false positive) rate in a 14AFC task, and is consistent with the fact that the range of possible target locations was relatively large, and infants did not tend to search actively for the target.

Adult Data

Adult acuity estimates with ACTIVE, measured at 10% contrast, were consistent with results reported previously for an 8° eccentric, 10% contrast, Gabor grating^{23,24} (Fig. 7). Thus, neither the thresholds of the first ($t_7 = -1.48$, $P = 0.183$, ns) or the second ($t_7 = -0.08$, $P = 0.936$, ns) test run differed significantly from the value of 7.7 cpd reported previously (see Fig. 2 of Abdelnour and Kalloniatis²⁵). There was some indication that group-mean acuity improved across the two runs. However, this difference was not significant (paired t -test; $t_7 = 0.97$, $P = 0.365$), and was largely caused by a single observer who performed initially near floor and improved by 3.5 octaves (see Fig. 7, black squares, for group-mean performance when this observer was excluded).

DISCUSSION AND CONCLUSIONS

The idea of using eye tracking to measure acuity is not a new one.^{44,45} Furthermore, semiautomated measures, in which the

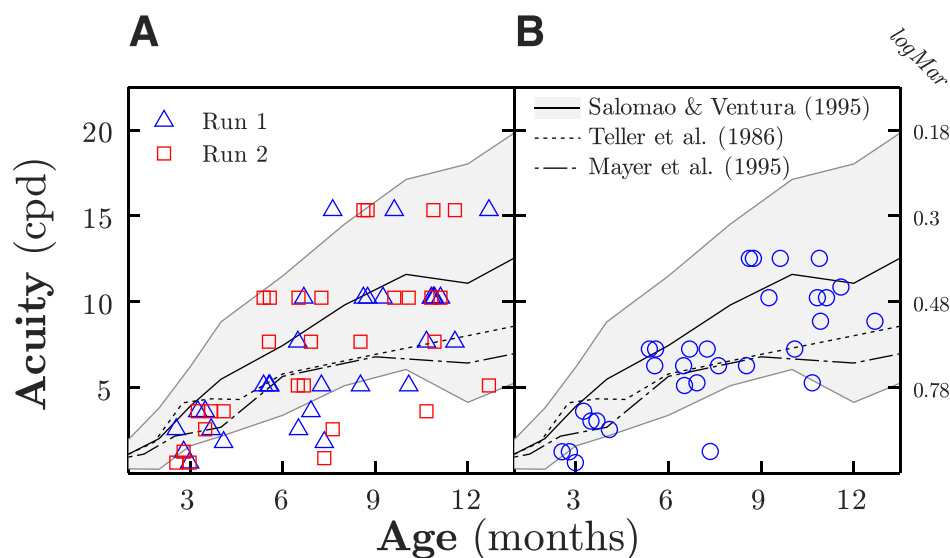


FIGURE 3. Visual acuity for individual infants, measured using the ACTIVE procedure. (A) Individual runs. (B) Geometric mean of two runs. For comparison, normative Teller acuity card data are given for: 206 infants tested binocularly by Teller et al.¹² (dotted line), 646 infants tested binocularly by Salomão and Ventura¹⁸ (solid line), and 269 infants tested monocularly by Mayer et al.²⁰ (1995; dash-dot line). The gray shaded region shows the 90% tolerance limits for Salomão and Ventura's¹⁸ data, within which 90% of the population would be expected to fall with 95% probability.

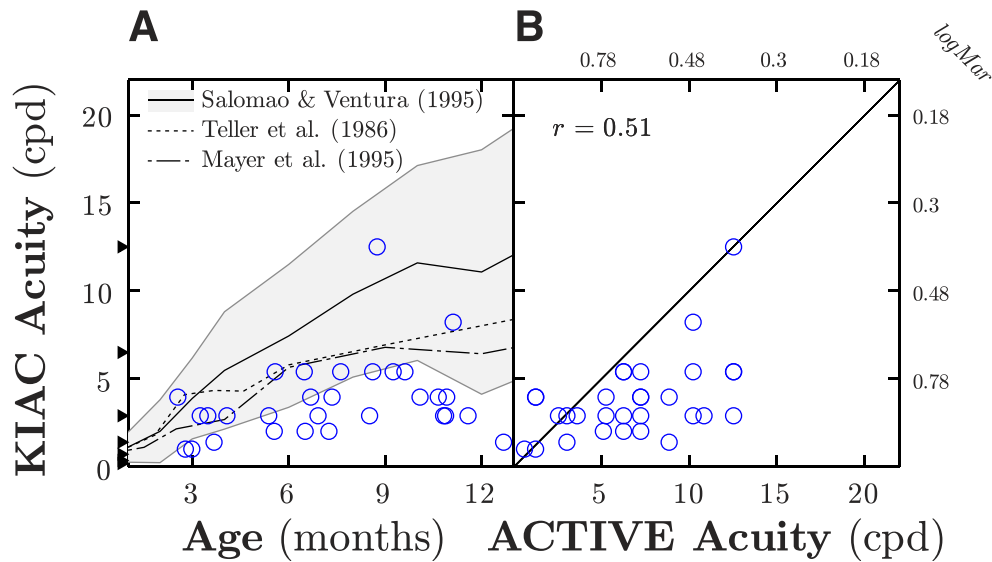


FIGURE 4. Geometric-mean visual acuity as measured using Keeler infant acuity cards, shown as a function of (A) age, and (B) each individual's corresponding eye-tracking estimate (see Fig. 3). Arrows on the ordinate show the available stimulus levels in the test set.

stimuli are generated by a computer but a human operator classifies the infant's responses, have been reported previously.^{1,2,46-49} However, the reported ACTIVE test is, to our knowledge, the first fully automated acuity measure shown to give reliable results in infants.

In terms of testability, test durations, and test-retest repeatability,^{19,20} the ACTIVE test performed comparably with the current gold standard (acuity cards). In terms of accuracy, the acuity estimates—especially when averaged over two runs—were consistent with normative data,^{12,18} and were correlated within-subjects with KIAC estimates. The estimates of KIAC were, however, consistently lower. Possible reasons for this are considered below.

Keeler Infant Acuity Card Estimates

Acuity estimates of KIAC were consistently lower than either the ACTIVE estimates, or previous normative data^{12,18} (derived principally using Teller cards). Because of this, previous Teller

card data were also used to validate the present ACTIVE procedure.

This disparity was unexpected and should be treated with caution. It is unlikely to have been because of the brand of acuity cards used per se, as extended sets of Keeler acuity cards have been shown to yield results similar to Teller acuity cards.³⁶ It is also unlikely to have been caused by differences in luminance, since the difference between the KIAC and ACTIVE tests was small in psychophysical terms,^{26,50} and luminance in the KIAC test was actually greater than in many previous studies.^{3,10,51}

One possibility is that acuity was underestimated in the KIAC test because of the relative inexperience of the two operators. A second possibility is that acuity was reduced because infant, parent, or operator moved backwards during testing (thereby making the stimuli harder to resolve). However, the most parsimonious explanation is that the low KIAC scores resulted from the spacing of the stimuli, which in the KIAC test vary from 0.94 to 1.16 octaves. This spacing is twice that of the acuity cards used in most normative studies,

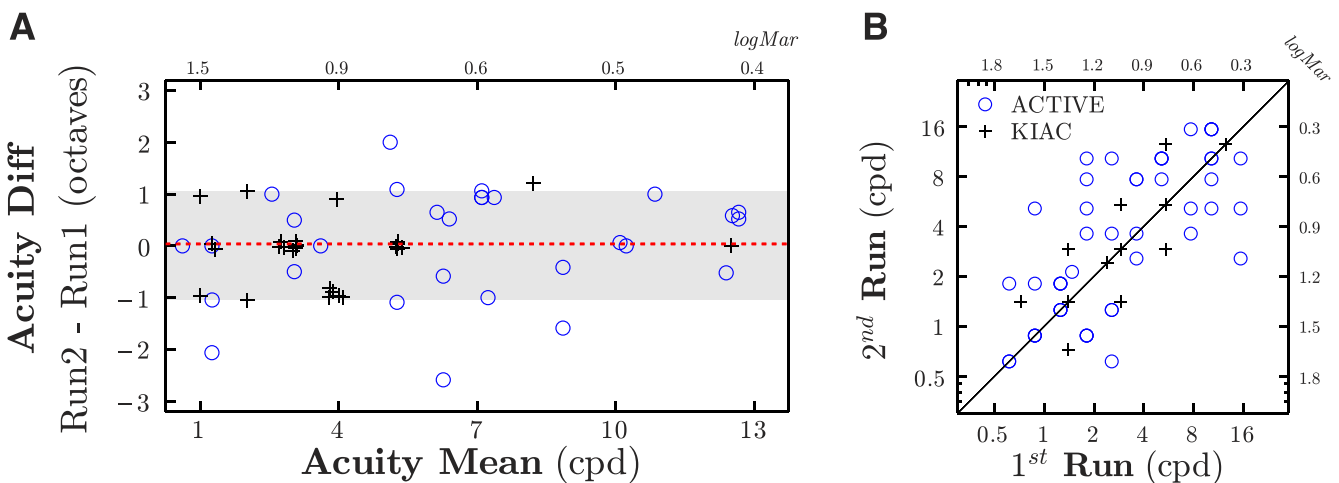


FIGURE 5. Test-retest reliability of the ACTIVE (circle) and KIAC (cross) procedures. (A) Bland-Altman plot showing how acuity estimates varied within subject, as a function of the mean. Overlapping points have been randomly jittered by a small amount for display purposes. (B) Scatter plot of the same data (without jitter).

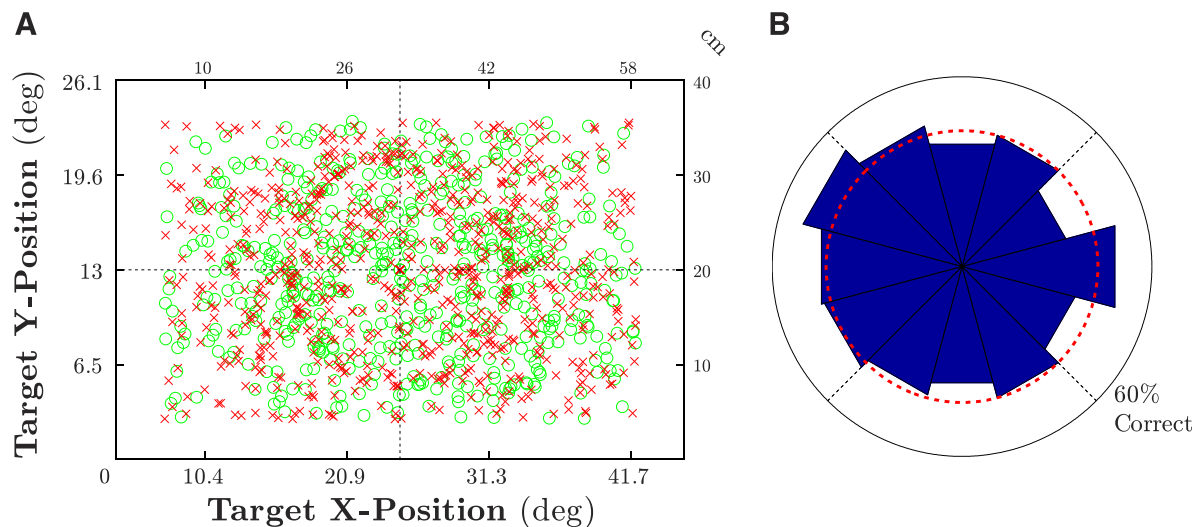


FIGURE 6. ACTIVE performance, as a function of (A) absolute target location (in degrees and centimeters, measured from the bottom-left of the screen), and (B) relative target location (relative to gaze-position at trial-onset). The green circles and red crosses in (A) indicate individual hit and miss trials, respectively, for all infants. The circular histogram in (B) indicates percent correct for 12 equal bins (n bins was arbitrary, but did not qualitatively affect the result), and the mean across all bins (red, dashed circle). Black dashed lines in (A) and (B) show the four location quadrants used during statistical analyses (see body text). The apparent annulus of points in (A); radius = 8°) reflects the fact that the probability of a central fixation at trial-onset was greater than chance (either because the previous target was located there, or because of the content of the intertrial cartoon).

and meant that no spatial frequencies were presented at the expected acuity of a 6- to 12-month-old (Fig. 4A, arrows). Consistent with this, we are aware of only one large scale study that has used full-octave (Teller) Acuity Cards in infants (Spierer et al.³), and that study also observed acuities that were substantially lower than those reported elsewhere. Without any conclusive evidence to support this interpretation, we can only tentatively suggest that full-octave acuity cards may not accurately estimate acuity, and that half-octave cards should be used where feasible.

Pros, Cons, and Related Measures

The adaptive computerized test of infant vision using eye tracking procedure allows resolution acuity to be measured automatically in normally sighted 3- to 12-month-olds, without the need for an experienced tester. Relative to acuity cards, it affords greater control over the stimulus in terms of its spatial-frequency content, luminance, and location in the visual field. Furthermore, by randomly varying the target location between trials, the chance of a false positive response is minimized (7%). The low guessing rate means that correct responses in the ACTIVE procedure are more informative than those in a traditional (e.g., 2AFC) paradigm, and helps ensure that the adaptive tracking algorithm converges on threshold quickly and robustly. The remaining advantages of the acuity cards are primarily their portability and initial cost (approximately five times cheaper given present hardware, though substantially cheaper eye-trackers are becoming increasingly prolific).

It is important to note that the ACTIVE procedure does not represent a like-for-like replacement for a clinician, and expert knowledge would still be required when dealing with an atypical or uncooperative infant. Moreover, it should be noted that other, nonbehavioral, techniques already exist with which to objectively assess visual acuity in infants.⁵²⁻⁵⁷ For example, the amplitude of the visual evoked potential elicited in response to a high contrast gratings correlates robustly with behavioral acuity measures, and may provide a more sensitive measure of primary sensory function.^{55,56} In contrast, ACTIVE provides a more rapid measure of functional acuity, which

could potentially be performed in ordinary clinics, and could be interleaved with other behavioral assessments.

Future Directions

The present paper demonstrates the validity of the ACTIVE procedure, but a larger sample will be required to provide normative data. It also remains to be seen how effective the ACTIVE method is in younger infants, or in infants with vision disorders where nystagmus or strabismus may complicate eye tracking. Finally, we are in the process of collecting monocular data using the ACTIVE procedure. Monocular testing is

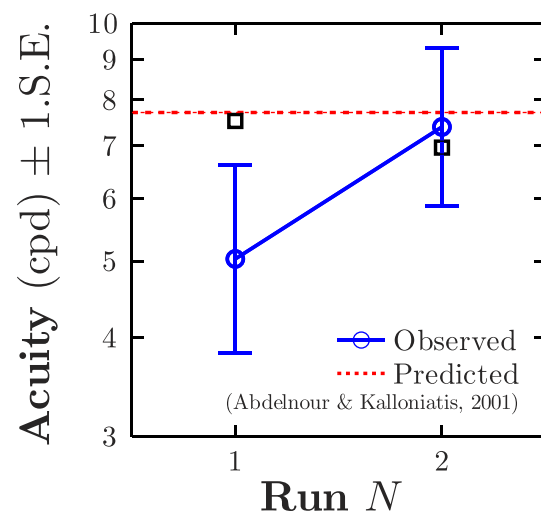


FIGURE 7. Adult group-mean acuity thresholds, tested at 10% contrast, as a function of two independent runs. The horizontal dashed line shows the predicted acuity given previously reported data. Black squares show group-mean performance when outlying data from a single observer were excluded (see body text). Means and standard-error were computed on the log-transformed data, and are displayed on a log-spaced y-axis.

particularly important for clinicians (e.g., when assessing amblyopia). Based on previous data,¹² we would expect monocular thresholds to be lower than those measured binocularly in the present work, by a factor of approximately $\sqrt{2}$.

Looking to the future, the most promising aspect of the ACTIVE procedure is that it can be readily modified to measure other visual functions, in addition to acuity. For example, by manipulating the luminance or hue of the target, sensitivity to contrast or color can also be assessed. Similarly, temporal modulation can be added to the stimulus to measure flicker-sensitivity, which can be diagnostic of nerve damage (e.g., in glaucoma⁵⁸) and retinal dystrophy.⁵⁹ Finally, the location of the stimulus relative to fixation can also be varied systematically, allowing assessment of the infant's visual field. These additional functions are crucial for fully characterizing the developing visual system, and methods of rapid assessment could be of substantial benefit in both clinical practice and research.

Acknowledgments

The authors thank John Wattam-Bell, Eliza Burton, Gary Rubin, and Michel Michaelides for their comments and suggestions.

Supported by Fight for Sight, the National Institute for Health Research Biomedical Research Centre at Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, the Special Trustees of Moorfields Eye Hospital, and the Leverhulme Trust.

Disclosure: **P.R. Jones**, None; **S. Kalwarowsky**, None; **J. Atkinson**, None; **O.J. Braddick**, None; **M. Nardini**, None

References

- Atkinson J, Braddick O, Bobier B, et al. Two infant vision screening programmes: prediction and prevention of strabismus and amblyopia from photo-and videorefractive screening. *Eye (Lond)*. 1996;10:189-198.
- Mercuri E, Haataja L, Guzzetta A, et al. Visual function in term infants with hypoxic-ischaemic insults: correlation with neurodevelopment at 2 years of age. *Arch Dis Childhood-Fetal Neonatal Ed*. 1999;80:F99-F104.
- Spierer A, Rozman Z, Chetrit A, Novikov I, Barkay A. Vision screening of preverbal children with Teller acuity cards. *Ophthalmology*. 1999;106:849-854.
- Hall HL, Courage ML, Adams RJ. The predictive utility of the Teller acuity cards for assessing visual outcome in children with preterm birth and associated perinatal risks. *Vision Res*. 2000;40:2067-2076.
- Mash C, Dobson V. Long-term reliability and predictive validity of the Teller Acuity Card procedure. *Vision Res*. 1998;38:619-626.
- Chandna A, Pearson CM, Doran RM. Preferential looking in clinical practice: a year's experience. *Eye (Lond)*. 1988;2:488-495.
- Atkinson J. *The Developing Visual Brain*. New York: Oxford University Press; 2002.
- Maurer D, Lewis TL, Brent HP, Levin AV. Rapid improvement in the acuity of infants after visual input. *Science*. 1999;286:108-110.
- Vaegan, Taylor DS. Critical period for deprivation amblyopia in children. *Trans Ophthalmol Soc U K*. 1978;99:432-439.
- McDonald MA, Dobson V, Sebris SL, Baitch L, Varner D, Teller DY. The acuity card procedure: a rapid test of infant acuity. *Invest Ophthalmol Vis Sci*. 1985;26:1158-1162.
- Fantz RL. Pattern vision in young infants. *Psychol Rec*. 1958;8:43-47.
- Teller DY, McDonald MA, Preston K, Sebris SL, Dobson V. Assessment of visual acuity in infants and children: the acuity card procedure. *Dev Med Child Neurol*. 1986;28:779-789.
- Fulton AB, Manning KA, Dobson V. A behavioral method for efficient screening of visual acuity in young infants. II. Clinical application. *Invest Ophthalmol Vis Sci*. 1978;17:1151-1157.
- Teller DY. The forced-choice preferential looking procedure: a psychophysical technique for use with human infants. *Infant Behav Dev*. 1979;2:135-153.
- Gwiazda J, Wolfe JM, Brill S, Mohindra I, Held R. Quick assessment of preferential looking acuity in infants. *Am J Optom Physiol Opt*. 1980;57:420-427.
- Fielder AR, Moseley MJ. Do we need to measure the vision of children? *J R Soc Med*. 1988;81:380-383.
- Courage ML, Adams RJ. Visual acuity assessment from birth to three years using the acuity card procedure: cross-sectional and longitudinal samples. *Optom Vis Sci*. 1990;67:713-718.
- Salomão SR, Ventura DF. Large sample population age norms for visual acuities obtained with Vistech-Teller acuity cards. *Invest Ophthalmol Vis Sci*. 1995;36:657-670.
- Mash C, Dobson V. Intraobserver reliability of the Teller acuity card procedure in infants with perinatal complications. *Optom Vis Sci*. 2005;82:817-822.
- Mayer DL, Beiser AS, Warner AF, Pratt EM, Raye KN, Lang JM. Monocular acuity norms for the Teller Acuity Cards between ages one month and four years. *Invest Ophthalmol Vis Sci*. 1995;36:671-685.
- Teller DY. First glances: the vision of infants. The Friedenwald lecture. *Invest Ophthalmol Vis Sci*. 1997;38:2183-2203.
- Tidbury LP, O'Connor AR. Testing vision testing: quantifying the effect of movement of visual acuity measurement [published online ahead of print October 24, 2014]. *Eye (Lond)*. doi:10.1038/eye.2014.247.
- Abdelnour O, Kalloniatis M. Word acuity threshold as a function of contrast and retinal eccentricity. *Optom Vis Sci*. 2001;78:914-919.
- Banks MS, Sekuler AB, Anderson SJ. Peripheral spatial vision: limits imposed by optics, photoreceptors, and receptor pooling. *J Opt Am A*. 1991;8:1775-1787.
- Dobson V, Teller DY. Visual acuity in human infants: a review and comparison of behavioral and electrophysiological studies. *Vision Res*. 1978;18:1469-1483.
- Dobson V, Salem D, Carson JB. Visual acuity in infants—the effect of variations in stimulus luminance within the photopic range. *Invest Ophthalmol Vis Sci*. 1983;24:519-522.
- Arden GB. The importance of measuring contrast sensitivity in cases of visual disturbance. *Br J Ophthalmol*. 1978;62:198-209.
- Jindra LF, Zemon V. Contrast sensitivity testing: a more complete assessment of vision. *J Cataract Refract Surg*. 1989;15:141-148.
- Johnson CA, Adams AJ, Casson EJ, Brandt JD. Blue-on-yellow perimetry can predict the development of glaucomatous visual field loss. *Arch Ophthalmol*. 1993;111:645-650.
- Johnson CA, Adams AJ, Casson EJ, Brandt JD. Progression of early glaucomatous visual field loss as detected by blue-on-yellow and standard white-on-white automated perimetry. *Arch Ophthalmol*. 1993;111:651-656.
- Johnson CA, Samuels SJ. Screening for glaucomatous visual field loss with frequency-doubling perimetry. *Invest Ophthalmol Vis Sci*. 1997;38:413-425.
- Cello KE, Nelson-Quigg JM, Johnson CA. Frequency doubling technology perimetry for detection of glaucomatous visual field loss. *Am J Ophthalmol*. 2000;129:314-322.
- Brainard DH. The psychophysics toolbox. *Spat Vis*. 1997;10:433-436.

34. Pelli DG. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis*. 1997;10:437-442.
35. Sireteanu R, Kellerer R, Boergen KP. The development of peripheral visual acuity in human infants. A preliminary study. *Hum Neurobiol*. 1983;3:81-85.
36. Neu B, Sireteanu R. Monocular acuity in preschool children: Assessment with the Teller and Keeler acuity cards in comparison to the C-test. *Strabismus*. 1997;5:185-202.
37. Clifford CE, Haynes BM, Dobson V. Are norms based on the original Teller Acuity Cards appropriate for use with the new Teller Acuity Cards II? *J AAPOS*. 2005;9:475-479.
38. Braddick O, Atkinson J, French J, Howland HC. A photorefractive study of infant accommodation. *Vision Res*. 1979;19:1319-1330.
39. Salapatek P, Bechtold AG, Bushnell EW. Infant visual acuity as a function of viewing distance. *Child Dev*. 1976;47:860-863.
40. Clifford-Donaldson CE, Haynes BM, Dobson V. Teller acuity card norms with and without use of a testing stage. *J AAPOS*. 2006;10:547-551.
41. Kaernbach C. Adaptive threshold estimation with unforced-choice tasks. *Percept Psychophys*. 2001;63:1377-1388.
42. Atkinson J, Braddick O. Visual attention in the first years: typical development and developmental disorders. *Dev Med Child Neurol*. 2012;54:589-595.
43. Holladay JT. Proper method for calculating average visual acuity. *J Refract Surg*. 1997;13:388-391.
44. Borchert MS, inventor; Children's Hospital Los Angeles, assignee. Methods and apparatus for measuring visual acuity in preverbal children. US patent US6402320. June 11, 2002.
45. Hathibelagal A. Objective assessment of visual acuity in infants. Presented at: XIV Biennial Meeting of the Child Vision Research Society; 2013; Waterloo, CA.
46. Shin YJ, Lee IB, Wee WR, Lee JH, Hwang JM. A novel computerized visual acuity test for children. *Korean J Ophthalmol*. 2013;27:194-198.
47. Atkinson J, Braddick O, Pimm-Smith E. "Preferential looking" for monocular and binocular acuity testing of infants. *Br J Ophthalmol*. 1982;66:264-268.
48. Lewis TL, Maurer D, Brent HP. The development of visual resolution in infants and toddlers tested monocularly with optokinetic nystagmus. *Clin Vis Sci*. 1990;5:231-241.
49. Lewis TL, Maurer D, Chung JYY, Holmes-Shannon R, Van Schaik CS. The development of symmetrical OKN in infants: quantification based on OKN acuity for nasalward versus temporalward motion. *Vision Res*. 2000;40:445-453.
50. Brown AM, Dobson V, Maier J. Visual acuity of human infants at scotopic, mesopic and photopic luminances. *Vision Res*. 1987;27:1845-1858.
51. Mayer DL, Dobson V. Assessment of vision in young children: a new operant approach yields estimates of acuity. *Invest Ophthalmol Vis Sci*. 1980;19:566-570.
52. Bane MC, Birch EE. VEP acuity, FPL acuity, and visual behavior of visually impaired children. *J Pediatr Ophthalmol Strabismus*. 1991;29:202-209.
53. Hamer RD, Norcia AM, Tyler CW, Hsu-Winges C. The development of monocular and binocular VEP acuity. *Vision Res*. 1989;29:397-408.
54. Norcia AM, Tyler CW. Spatial frequency sweep VEP: visual acuity during the first year of life. *Vision Res*. 1985;25:1399-1408.
55. Sokol S, Moskowitz A. Comparison of pattern VEPs and preferential-looking behavior in 3-month-old infants. *Invest Ophthalmol Vis Sci*. 1985;26:359-365.
56. Riddell PM, Ladenheim B, Mast J, Catalano T, Nobile R, Hainline L. Comparison of measures of visual acuity in infants: Teller acuity cards and sweep visual evoked potentials. *Optom Vis Sci*. 1997;74:702-707.
57. Suttle, CM. Visual acuity assessment in infants and young children. *Clin Exp Optom*. 2001;84:337-345.
58. Tyler CW. Specific deficits of flicker sensitivity in glaucoma and ocular hypertension. *Invest Ophthalmol Vis Sci*. 1981;20:204-212.
59. Tyler CW, Ernst W, Lyness AL. Photopic flicker sensitivity losses in simplex and multiplex retinitis pigmentosa. *Invest Ophthalmol Vis Sci*. 1984;25:1035-1042.