1040-5488/15/9207-0823/0 VOL. 92, NO. 7, PP. 823–833
OPTOMETRY AND VISION SCIENCE

ORIGINAL ARTICLE

Measuring Infant Visual Acuity with Gaze Tracker Monitored Visual Fixation

Amithavikram R. Hathibelagal*, Susan J. Leat†, Elizabeth L. Irving‡, Krithika Nandakumar§, and Moshe Eizenman‖

ABSTRACT

Purpose. To validate a method of measuring grating acuity with remote gaze tracking (GT) against a current clinical test of visual acuity (VA), the Teller Acuity Cards (TACs), as part of the development of an automated VA test for infants.

Methods. Visual acuity for computer-generated horizontal square-wave gratings was determined from relative fixation time on a grating area compared with the background. In experiment 1, binocular VA was based on eye movements with a GT in 15 uncorrected myopic adults and compared with VA measured with subjective responses with the same stimuli and with the TACs. In experiment 2, binocular VA was determined in 19 typically developing infants aged 3 to 11 months on two visits with both the GT and TACs.

Results. In adults, the mean difference between VA measured by the GT and TACs was 0.01 log cycles per degree (cpd) and the 95% limits of agreement were 0.11. One hundred percent of GT VA results were within 0.5 octave of the TACs' VAs. The mean difference between the GT and TACs for infants was 0.17 log cpd on both the first and second visit (95% limits of agreement, 0.42 and 0.47, respectively). The mean difference between test and retest for infant GT VA was 0.06 log cpd, and limits of agreement for repeatability were 0.48 log cpd. In infants, both the TACs and the GT had a reliability of 89% within less than or equal to 1 octave between visits. Gaze tracking VA improved with age and is in agreement with published norms.

Conclusions. The agreement between the TACs and GT in adults and infants validates the method of measuring grating acuity with the remote GT. These results demonstrate its potential for an automated test of infant VA.
(Optom Vis Sci 2015;92:823–833)

Key Words: gaze tracker, grating acuity, eye movements, visual fixation, infant vision, preferential looking

Early detection of abnormal visual acuity (VA) is crucial in the identification and management of ocular and visual abnormalities in infants.[1] Currently, Teller Acuity Cards (TACs) are considered the gold standard for clinical testing[2] and are effective in obtaining a quick estimate of an infant's VA,[3] but they have drawbacks.[4] They rely on a subjective assessment of the infant or child's looking behavior including a qualitative assessment of eye movements, head movements, facial expressions, pointing, or verbal responses. These parameters are not quantitative and may lack objectivity.[5] To minimize observer bias, the TACs can be used with a strict two-alternative forced-choice paradigm, so that the observer is unaware of the grating position and the absolute spatial frequency. Clinically, however, the test is more often performed using the acuity card procedure in which the tester is not masked to the actual spatial frequency, and this may lead to bias.[6,7] Despite this, TACs have been found to have good validity and reliability.[8] In a previous paper,[9] we demonstrated the ability to use a gaze tracking (GT) system to measure grating VA objectively in adults by analyzing visual scanning patterns for targets of different spatial frequencies. In the testing protocol with adults, the spatial frequency of the target changed over a large range of spatial frequencies (1.5 to 35.1 cycles per degree [cpd]). The adult participants looked at suprathreshold gratings for about 72.5% of the total time (5 seconds) that the gratings were presented. Infants' attention span is limited, precluding use of the same protocol. The current proof-of-concept study instead investigates the use of a similar method with a modified VA test protocol. A human observer assesses the information generated by the Gaze Tracker (GT) to arrive at a measure of VA. The data from the study will be useful in determining whether a fully automated system to measure grating VA in infants is feasible.

*BOptom, MSc
†PhD, FCOptom, FAAO
‡OD, PhD
§BOptom, PhD
‖PhD
School of Optometry and Vision Science, University of Waterloo, Ontario, Canada (ARH, SJL, ELI, KN); and Institute of Biomaterials and Biomedical Engineering, University of Toronto, Ontario, Canada (ME).

This article describes two experiments. Experiment 1 developed and validated a GT VA estimation protocol in naïve adults with uncorrected refractive error. Experiment 2 adapted the protocol developed in experiment 1 to measure GT VA in infants and to determine the testability, test-retest repeatability, and validity of the GT. Validity in infants was determined by comparison with TACs and age-related published normal values. We hypothesized that the testability and repeatability of the GT would be at least as good as that of TACs and that the VA obtained with the GT would be comparable to the VA obtained by TACs.

METHODS

Remote Noncontact GT

The remote gaze estimation system used in this study extracts eye features from video images and uses these features to estimate gaze position.[10–12] The GT system estimates the direction of the optical axes of the two eyes without user calibration and uses a probabilistic calibration approach that does not require continuous fixation on targets to estimate angle κ (angle between the visual and optical axis). During calibration, small targets (cartoon characters with sound) appeared at random positions on the display monitor.[13] The system requires very limited participant cooperation and can be used with infants.[13] The remote GT system consists of two 21-inch LCD monitors, 7 infrared light sources, and 2 video cameras (Point Grey Grasshopper 20S4M-C) (Fig. 1). We used only three infrared light sources to test each infant, but the ability to select a specific set of lights and to optimize the illumination pattern for an infant increased the robustness of the system. The two video cameras were located 70 cm from the participant and below his or her line of sight. The GT has an accuracy of ±0.5 degrees and a spatial resolution of less than 0.1 degrees. The same instrument was used for both adult and infant studies. One of the monitors (monitor 2; Fig. 1) displayed the visual stimulus (horizontal square-wave gratings) that could be alternated with cartoon videos between stimuli to maintain the infant's attention. This monitor was movable so that the distance of the visual stimuli from the participant could be adjusted. Horizontal gratings that filled rectangular areas on the monitor's screen were presented randomly in one of four alternative positions on the monitor. The gratings had equal luminance to the background, so that the grating patches were invisible to a participant if he or she could not resolve the grating. The use of four alternate positions decreased the possibility that a looking response was judged to be correct by chance. The other monitor (monitor 1; Fig. 1) allowed the experimenter to control the procedures and displayed the eye tracking data in real time.

The eye-tracking information allowed an observer to judge if the participant could resolve the grating. The information showed the real-time gaze position of each eye (clusters of white and gray dots [red and blue dots in the online version] in Fig. 2A, C) relative to the position of the grating, the elapsed time from the start of the stimulus presentation, and a bar graph (Fig. 2B, D), giving the percentage of time spent fixating the grating (gray [green online]) compared with the time spent on other regions of the monitor (white).

Dimmed room lights for GT testing decreased distractions from the surroundings. The mean luminance of the GT gratings was 67 cd/m^2 (Minolta Chroma Meter CS 100 photometer). The TAC test was performed in general room lighting.[14] The mean luminance of the TACs was 45 cd/m^2.

Testing distance for adults for both TACs and the GT was 210 cm, where spatial frequencies for the GT ranged from 0.58 to 37 cpd, in about 0.5-octave steps. The testing distance for infants

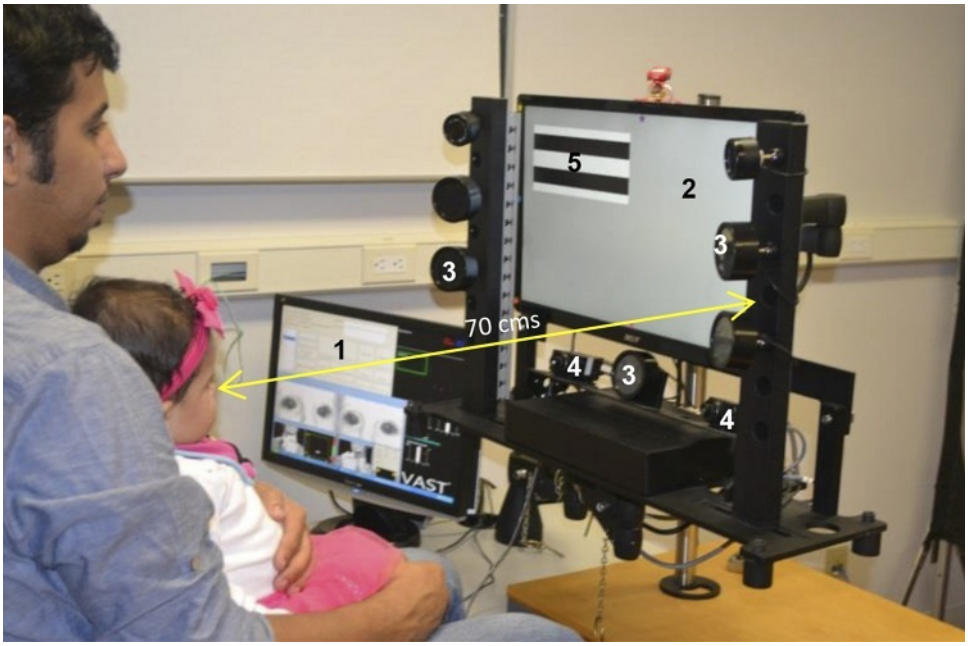

FIGURE 1.
Infant participant with the GT. (1) Monitor 1, (2) monitor 2, (3) infrared light sources (seven), (4) two video cameras, (5) grating target/stimulus. Note that during the actual testing, there was a screen between the two monitors, so that the infant would not be distracted by monitor 1. A color version of this figure is available online at www.optvissci.com.

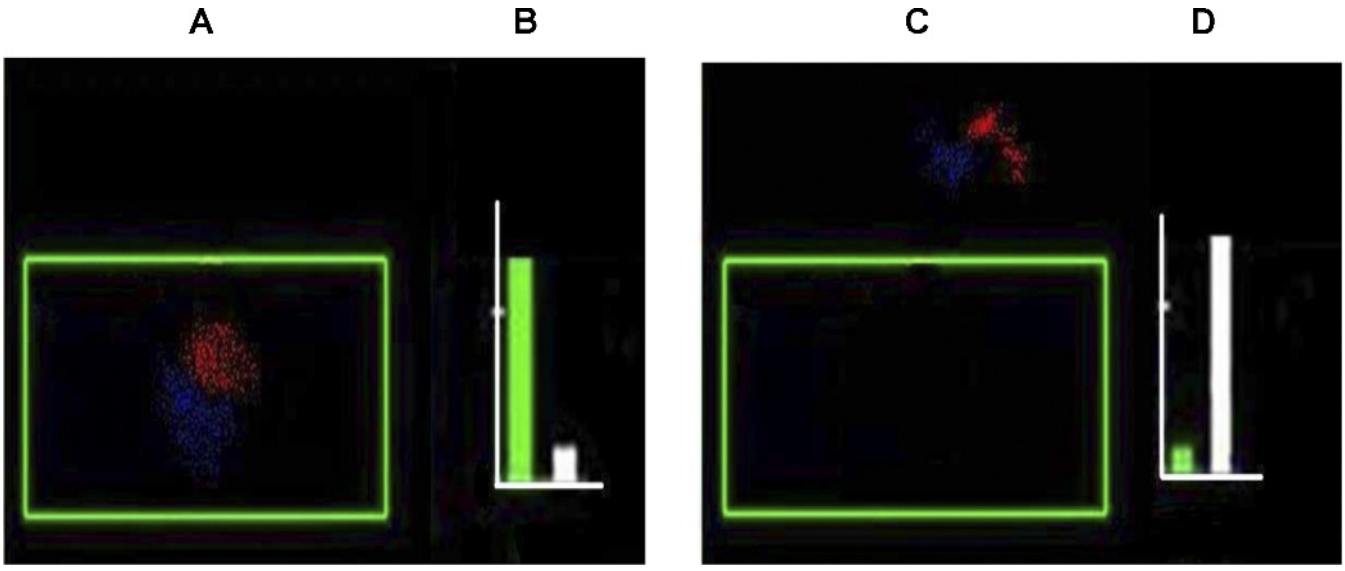

FIGURE 2.

Clipped regions of the experimenter monitor used to make seen/unseen judgments. In A and C, the white and gray cluster of dots (red and blue online) represent the left eye and right eye gaze positions, respectively (the display shows the latest 3 seconds of data, and the data for the left eye is shifted by 1 degree horizontally and vertically relative to the right eye so that the eye gaze positions of the two eyes do not overlap, as viewed by the observer). The white rectangular outline (green online) represents the grating target area. In B and D, the gray bar (green online) displays the time spent fixating within the grating and the white bar represents the time spent on the rest of the monitor. In A and B, the participant was fixating on the gratings for more than 75% of the time (the gray bar [green online] in B is above the marker on the vertical axis), which was judged as seen. In C and D, the participant was not fixating on the grating target for either 75% of the time or 2 continuous seconds within 10 seconds, so this was judged as unseen. A color version of this figure is available online at www.optvissci.com.

was either 70 or 120 cm depending on age; this closer distance enabled spatial frequencies low enough for younger infants. Infants 6 months or younger were tested at 70 cm, and the horizontal grating spatial frequencies ranged from 0.2 to 12.5 cpd in 0.5-octave steps. Infants older than 6 months were tested at 120 cm, and the horizontal spatial frequencies ranged from 0.34 to 21.4 cpd. Although every effort was made to have the same spatial frequencies for the TAC and the GT, there were some slight discrepancies. Why? The TACs were not exact 0.5-octave steps, and the discrete pixels on the computer screen precluded exact 0.5-octave steps (for that, the grating stripe width needed to be exact multiples of 1.41). The overall size of the rectangular targets subtended at least 17 × 6.5 degrees at the testing distance of 70 cm, 9.3 × 3.8 degrees at 120 cm, and 2.1 × 1 degree at 210 cm.

Protocols and methods were suitable for infants, but first validated in adults. During the test, the observer determined whether the gratings could be resolved by a participant by using data that were provided in real time by the GT. A "seen" judgment was made when one or both of the following criteria were met:

1. If the initial fixation was within the grating target and the participant fixated at least 2 seconds within the target area without his or her fixation leaving the target area. This tended to happen with gratings that were well above threshold. If this did not occur, the targets were judged as "seen" if there was a sequence of fixations for at least 2 seconds within the target area, within a presentation window of 10 seconds for adults or 6 seconds for infants. The difference between adults and infants was based on the observation that adults appeared to search for the target when it was close to their VA threshold so a longer search window was beneficial, whereas a pilot study with a different group of infants showed that infants tended to lose attention during longer presentations and were less likely to search for the target. See Hathibelagal[15] for details. For adults, the grating target was less than 8% of the monitor's area and the requirement that the participant spend at least 2 out of 10 seconds (20%) within the target boundaries limited the probability of false positives. The expected fixation time by chance, when the gratings were below threshold, was 8%. In the infant study, the gratings occupied a maximum of 19% of the monitor's area. This was less than the criterion level (2 out of 6 seconds = 33%).
2. If the bar chart, which indicated the ratio of the total fixation time within the grating target over the total monitor surface, reached 75%. Again, this criterion was significantly above chance for both adults and infants.

The tenets of the Declaration of Helsinki were followed, and the study was reviewed and obtained clearance through the Office of Research Ethics at the University of Waterloo.

Experiment 1: Adults

Participants

Adults (graduate students at the University of Waterloo, Canada) provide much more data in one sitting than infants, thus allowing a comparison of a greater number of protocols. Although the agreement between different tests would be anticipated to be better in adults, with less variability, we considered that a test or protocol that did not show good validity in adults would be highly unlikely to do so in infants.

Sample size calculation was based on the results of McDonald et al.[16] who found an intraobserver correlation of 0.66 for the acuity card procedure in infants. We chose to use published data from preferential acuity determination, rather than subjective measures, as the purpose of the present study was to develop a method based on fixation patterns. To obtain a significant correlation for 80% power and an α level of 0.05, 15 to 16 participants were required. Inclusion criteria were as follows: corrected VA of at least 20/20 and myopia with a spherical equivalent between −1.25 diopters (D) and −10 D. Uncorrected myopes were used to obtain a range of VA and to ensure that fluctuating accommodation would not influence the results. Exclusion criteria included a history of ocular disorders other than myopia and astigmatism, strabismus, previous ocular surgery, or astigmatism greater than 2.50 D (any axis).

Psychophysical Procedures Used for Adults

The order of testing for the adults was GT followed by TACs. The same testing distance of 210 cm was used for all measures of VA in adults. The same person acted as experimenter (controlling the experiment) and observer, making the "seen" or "unseen" judgments for both the GT and TACs.

Eye Movement–Based GT Protocol

The adult participants were naïve for the GT measures; that is, they were asked to look at the monitor, but given no further instructions. This was implemented to imitate the situation with infants and rely on natural eye movements. This technique was called eye movement–based GT VA in adults ("GT").

A number of preliminary protocols for the GT acuity measurement were compared (details described by Hathibelagal[15]). The protocols gave the same acuity; thus, the most time-efficient protocol was chosen. This was a staircase starting with one to four presentations at a low spatial frequency, increasing in 1-octave steps. After the first reversal, that is, when a "seen" judgment was not made, gratings were presented at every 0.5-octave level. The number of presentations, up to a maximum of 4, was at the discretion of the experimenter, who was allowed to repeat presentations until he was satisfied the grating was seen. At spatial frequencies that were well above threshold, fixation responses were immediate and obvious; two or less presentations were needed at each spatial frequency at the beginning of the staircase. As spatial frequency increased, the fixation responses were less immediate, and the final VA criterion was three "seen" judgments out of four presentations at one level with at least two incorrect judgments in the next higher spatial frequency. The "seen" judgment was based on the fixation patterns and the histogram information, according to the two criteria described. This protocol is similar to that used clinically for TACs.

Teller Acuity Cards

The participant's VA was determined using TACs at 210 cm. The acuity was measured without instructions, relying on fixations, as with an infant. Teller Acuity Cards were held in the usual orientation (card held horizontally, vertical gratings) to compare the GT VA with TACs norms. Teller Acuity Cards were also held in a vertical orientation, with gratings horizontal. This is not typical for TACs but was included so that the grating orientation matched that of the eye tracker. It would be expected that there would be more concordance of acuity between the GT protocol and the TAC when the gratings have the same orientation, especially in cases of high astigmatism in which the blur would be different in the two meridians.

Analysis

Visual acuity was analyzed in terms of log cpd. The D'Agostino-Pearson omnibus test determined the normality of the VA data. Agreement was assessed with the 95% limits of agreement ($\pm 1.96 \times$ the SD of the difference between tests).[17] The percentage of participants' VA results, which agreed between methods, was calculated (i.e., the percentage of participants whose results [e.g., comparing GT and TACs] were different by $\leq$0.5 plus 5% octave or $\leq$1 plus 5% octave, as is commonly used).[18] The 5% additional margin was allowed because the spatial frequencies of the two tests were not identical. A repeated-measures analysis of variance (ANOVA) investigated differences in VAs between the GT and TACs for horizontal and vertical orientations. Analyses used Excel, Statistica (StatSoft Corp, USA), and Graph Pad Prism.

RESULTS

Fifteen adults (six women and nine men) ranging from 22 to 47 years (mean, 28.47 years) participated. Visual acuity was normally distributed in all protocols ($p < 0.05$), except one, which was borderline ($p = 0.05$); hence, parametric tests were used. Repeated-measures ANOVA ($F_{3,42} = 2.44$, $p = 0.08$) showed no significant statistical difference between the mean thresholds obtained from the eye movement–based GT protocol and horizontal and vertical orientations of TACs.

The agreement between TACs and GT acuities is shown in the Bland-Altman plot in Fig. 3. The mean difference between the GT and the TACs (horizontal gratings, GT − TACs) was 0.01 log cpd, and the 95% limits of agreement were ±0.11. It can be seen in Fig. 3 that all the GT data points were within 0.5 octaves of the TACs (horizontal gratings). With the TACs held in the usual orientation (vertical gratings), all TACs VA results were within 0.5 octave of the GT result. Tables 1 and 2 summarize these results for adults.

Experiment 2: Infants

Participants

Participants were recruited from the Grand River Hospital Birth Clinic, local bulletin boards, and doctors' offices and through patient records available at the Pediatric and Special Needs Clinic at the University of Waterloo School of Optometry and Vision Science. Eligibility criteria were: (1) age between 3 and 12 months, (2) gestational age between 37 and 42 weeks at birth, (3) no known major medical problems, (4) normal development by parental report and observation, and (5) completion of basic eye screening examination, which included Hirschberg and unilateral cover test to check for the presence of strabismus, broad H to test for incomittancy, and refractive error measured by

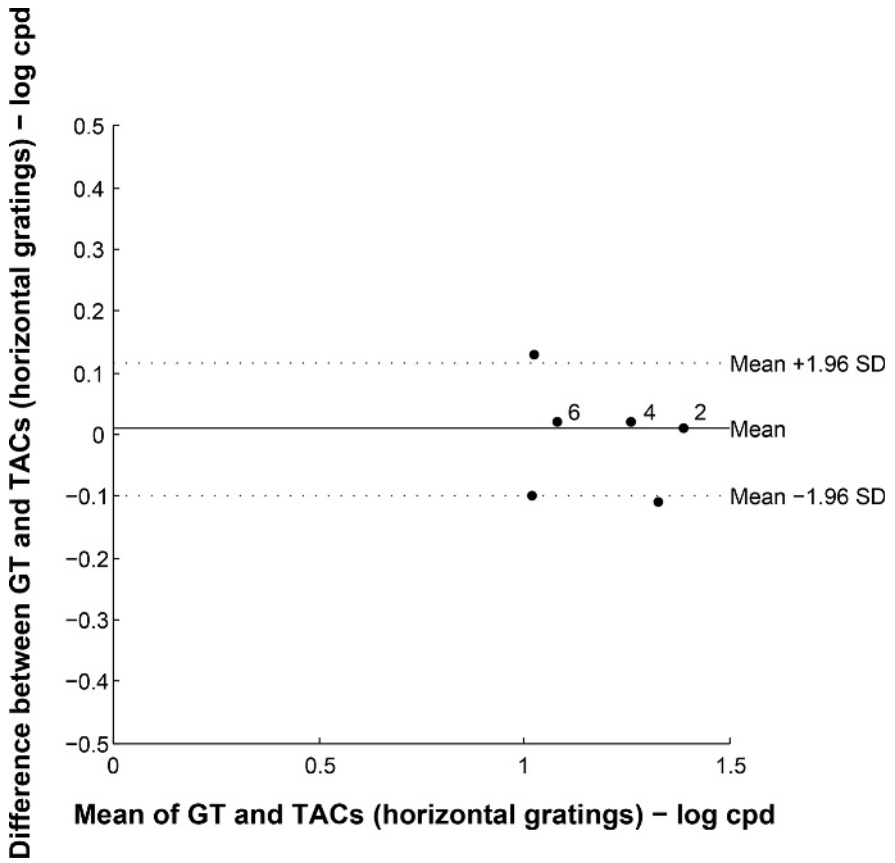

FIGURE 3.

Agreement between eye movement–based GT acuities and TACs acuities (horizontal gratings) for adult participants. The difference between the acuities is plotted against the mean of the acuities. The digits indicate the number of data points where there is more than one data point in a given location. The dotted lines indicate the 95% limits of agreement.

Mohindra retinoscopy. Spherical refractive error exclusion criteria were based on the noncycloplegic retinoscopy results of Gwiazda et al.[19] for normal infants and were as follows: myopia exceeding −2.00 and −1.50 D for infants 6 months and younger and those older than 6 months, respectively, and hyperopia exceeding +4.00 and +3.00 D for infants 6 months and younger and infants older than 6 months old, respectively.[19] Infants with astigmatism greater than 2.50 D at any axis were excluded.[20] We aimed for a sample size of 20 to ensure that we were not underpowered because of the greater variability of infant's responses.

Psychophysical Procedures Used for Infants

Visual acuity was measured with the GT and TACs on two occasions to determine the validity (comparison of GT with TACs) and test-retest repeatability of both tests. The two sessions were scheduled within 7 days for infants aged 6 months or younger and within 10 days for infants older than 6 months.[8]

All the acuity observations (TACs and GT) were made by a single, experienced observer, so as not to confound the results with interobserver differences. The experienced observer was a pediatric optometrist, who had previous experience in acuity testing in infants. The order of testing (GT or TACs first) was randomized between participants, with the constraint that an equal number of participants started with the GT and with the TACs. The same order was used for the second visit.

Gaze Tracker

For the GT, the infant was seated on the parent's lap on an adjustable chair. The parent's eyes were not visible to the observer and were not tracked by the GT. Monitor 1 was not visible to the infant as a screen was placed between the two monitors. For infants, two people were involved in testing, one taking the role of experimenter and the other taking the role of observer. The experimenter controlled the order of spatial frequencies in the

TABLE 1.

Means (±SD) of VA in log cpd for GT and TACs in adults and infants

Adults	**GT**	**TACs (horizontal gratings)**	**TACs (vertical gratings)**	
	1.18 ± 0.13	1.17 ± 0.14	1.21 ± 0.13	
Infants	**GT visit 1**	**GT visit 2**	**TACs visit 1**	**TACs visit 1**
	0.89 ± 0.30	0.83 ± 0.30	0.73 ± 0.22	0.67 ± 0.24

TABLE 2.

Agreement between GT and TACs (horizontal gratings) in adults and infants is shown together with repeatability statistics (test-retest) for infant study

Adult study—agreement between TACs and GT		
Percent agreement ≤ 0.5 octave	100%	
Percent agreement ≤ 1 octave or less	100%	
Mean difference (GT − TACs)	0.01	
95% limits of agreement	±0.11	
Infant study—agreement between TACs and GT		
	First visit	**Second visit**
Percent agreement ≤ 0.5 octave	63% (12/19)	47% (9/19)
Percent agreement ≤ 1 octave	74% (14/19)	58% (11/19)
Mean difference (GT − TACs)	0.17 log cpd	0.17 log cpd
95% limits of agreement	±0.42 log cpd	±0.47 log cpd
Infant study—repeatability		
	GT	**TACs**
Percent agreement ≤ 0.5 octave	63% (12/19)	47% (9/19)
Percent agreement ≤ one octave	89% (17/19)	89% (17/19)
Average difference (1st − 2nd visit)	0.06 log cpd	0.06 log cpd
95% limits of agreement	±0.48 log cpd	±0.38 log cpd

protocol (based on the judgments from the observer) and also determined the starting spatial frequency for each participant. The observer tapped on the experimenter's right or left shoulder to indicate a "seen" or "unseen" judgment, respectively, and in the center of his back for a presentation that "needs repeating." The observer was masked regarding the actual spatial frequencies to reduce bias.

The protocol chosen for the infant study was similar to the adult GT protocol, except that it was a 1-down, 1-up staircase with the step size changing from 1 octave to 0.5 octave at the first reversal (unseen judgment). After the third reversal, a criterion of at least three correct out of four presentations was applied before moving up or down in spatial frequency. The infant's attention was maintained by showing cartoon videos between the stimuli.

Teller Acuity Cards

The observer knew each participant's name but not the starting spatial frequency, participant's age, spatial frequencies of the gratings, final TAC acuity, or the acuity obtained by the GT if assessed first. The choice of the low spatial frequency start card was randomized by a second person before the participant arrived, and the same start card was used for the second session for each infant. The start card used for each infant was numbered 1 and the next 0.5-octave step card was numbered 2 and this continued with sequential numbering to the highest spatial frequency. The TACs were used in a vertical orientation, so that the gratings were horizontal. All testing was done with the acuity cards placed behind a TAC stage. A TAC stage with three panels was constructed with a vertical opening in the central panel, in which the TACs were presented. The dimensions of the TAC stage were similar to that described by Clifford-Donaldson et al.[21] The background gray matched the background of the TACs. Infants were on the parent's lap for testing. The test distance was 55 cm for all the infants.

First, the observer showed a blank card to see the infant's response when the stimulus could not be resolved. Next, the observer presented the lowest spatial frequency, to determine how a "clear look" appeared. Then, she tested in 1-octave steps, starting at the start card and showing one card at each level as long as a clear, correct look was observed. The observer stated her judgment of the grating position to the experimenter who indicated whether her judgment was correct or not. Once a less clear look or an incorrect judgment was made, testing was done in 0.5-octave steps. Previously presented cards could be tested again at the discretion of the observer. Testing was considered complete when the observer was satisfied that a test card of the highest spatial frequency to give a clear, correct look had been obtained. To check, the observer always presented a card at least twice at the next higher spatial frequency (at which the grating was not resolved) than the estimated acuity.

Analysis

The analysis of the infant data included agreement and repeatability between TACs and the GT analyzed with Bland-Altman plots and calculation of the 95% limits of agreement. Additionally, the percentage agreement and repeatability within 0.5 and 1 octave was tabulated. Repeated-measures ANOVA was used to assess the order or method on time taken and acuity obtained.

RESULTS

Of the 25 infant candidate participants, 5 were excluded based on refractive error inclusion criteria. Of the 20 infants, 55% (11) were female.

There was 100% completion for TACs on both sessions, whereas the GT had 100% completion rates for the first session and 95% for the second session (one 5-month-old did not finish the second GT protocol). Subsequent analysis was undertaken with the 19 participants, aged 3.2 to 11 months (mean [±SD] age, 8.1 [±2.43] months), who completed both tests. Three were between 3.2 and 3.5 months, and 16 were older than 6 months (mean [±SD] age, 9 [±1.34] months).

The VA data were normally distributed ($p > 0.05$). The means and SDs of the GT and TAC acuity for the two visits are shown in Table 1. The repeatability of the GT is shown in Fig. 4 and summarized in Table 2. The mean difference (first − second visit) was 0.06 log cpd and the limits of agreement were ±0.48 log cpd. The percentage agreement for repeated measures to less than or equal to 0.5 octave was 63% and 84% within 0.5 octave for the GT and TACs respectively. The percent agreement for repeatability less than or equal to 1 octave for both methods was 89%.

Agreement between the GT and TAC for the first and second visit is shown in Fig. 5. The limits of agreement (GT − TACs) were ±0.42 and ±0.47 log cpd for the first and second visit, respectively (Table 2). Seventy-four percent and 58% of measures were within less than or equal to 1 octave of each other for the first and second visits, respectively, as shown in Table 2. Repeated-measures ANOVA (2 visits × 2 methods) showed a main effect of method ($F_{1,18} = 14.8$, $p = 0.001$), but there was no main effect of visit ($F_{1,18} = 1.94$, $p = 0.18$). There was no interaction between method and visit for the acuity obtained ($F_{1,18} = 0.001$, $p = 0.98$). The GT estimates were, on average, higher than TACs in both visits (mean difference, 0.17 log cpd for both visits). Validity was also investigated by plotting VA against age. There were significant correlations between GT acuities and age ($r = 0.80$ and 0.73,

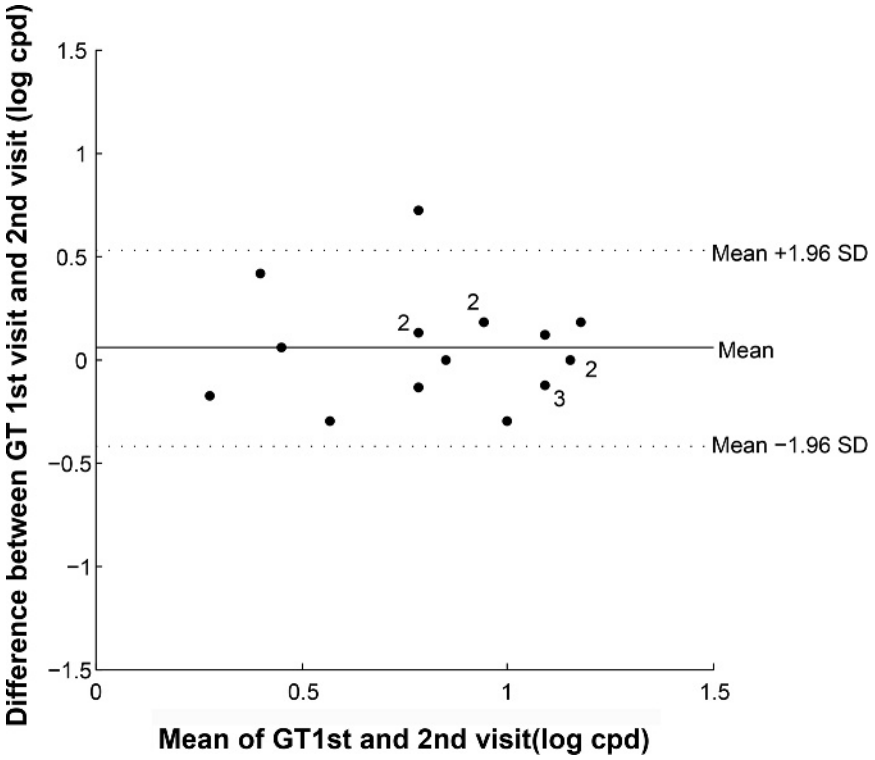

FIGURE 4.

Repeatability of VA obtained with the GT over two visits for infants. The difference between the test and retest of the acuities is plotted against the mean of the two acuities. The digits indicate the number of data points where there is more than one data point in a given location. The dotted lines indicate the 95% limits of agreement.

$p < 0.05$, for visits 1 and 2, respectively). These correlations were not significantly different from each other ($p = 0.63$). There were also significant correlations between TACs acuities and age ($r = 0.58$, $p < 0.05$; $r = 0.45$, $p = 0.05$ for the first visit and second visit, respectively). The average acuities of the two visits for each infant for both GT and TACs with respect to age are plotted in Fig. 6 together with the 95% range from Salomao and Ventura[22] and Courage and Adams.[5]

The mean (±SD) for time taken for the first and second visit for TACs was 5.2 (±1.4) and 4.5 (±0.88) minutes, respectively, and for the GT, it was 5.6 (±1.1) and 5.6 (±1.43) minutes, respectively. These times do not include the setup time for either test, that is, seating the child at the right height and, in the case of the GT, undertaking the calibration. Repeated-measures ANOVA for the time taken (2 visits × 2 methods) showed a main effect of method ($F_{1,18} = 6.77$, $p = 0.02$), the TAC method being significantly faster than the GT method. There was no main effect of visit ($F_{1,18} = 1.3$, $p = 0.27$) and no interaction between visit and method ($F_{1,18} = 2.5$, $p = 0.13$).

DISCUSSION

Adult participants in experiment 1 developed the protocol and determined potential validity of the GT for use with infants. Adults showed 100% agreement between eye movement–based GT and TACs (horizontal gratings) within 0.5 octave, and the 95% limits of agreement were ±0.11 log cpd. This is similar to limits of agreement for repeatability for VA measured with the same letter chart in adults, which is of the order of one line or 0.1 logMAR (logarithm of the minimum angle of resolution).[23] The agreement in the current study was better than the agreement between different charts of VA in young adults in a recent study comparing the logMAR Sloan letters with Landolt C's and tumbling E's in which the limits of agreement ranged between 0.18 and 0.12 logMAR.[24] Considering that the current study used observed rather than subjective responses, this confirms good validity for the GT against TACs. Protocol order in adults was not randomized, which may have introduced fatigue or practice effects. The similar thresholds suggest that this did not occur.

Experiment 2 showed good validation for the GT in infants based on comparison with the "gold" standard, that is, the TACs, and comparison with infant binocular acuity norms demonstrating an increase of VA with age. The agreement in the present study between GT and TACs acuities less than or equal to 0.5 octave and less than or equal to 1 octave is similar to agreement between different measures of infant binocular acuity (44 to 67% within 0.5 octave[18] and 66 to 100% within 1 octave).[6,16,18,25,26]

Fig. 6 shows that our results for both the TACs and the GT are within the reported reference ranges, with the exception of one point. It can also be seen that there is some difference in the reported norms, the data of Salomao and Ventura[22] being higher on average than that of Courage and Adams.[5] Our TAC data are more consistent with Courage and Adams than with Salomao and Ventura. The GT VAs cluster within both normative data sets (a couple of points are very slightly above the upper limit of the Courage and Adams data and one point is below the lower limit of the Salomao and Ventura data). Again, this shows good validity for the GT, that is, similar values and improvement with age as other studies.

Our GT results gave significantly higher VA than the TACs, especially for the higher VA values (Fig. 5). Perhaps, GT demands a smaller saccade to fixate on the grating. The edge of the gratings is equally close or closer to the center than TACs in the horizontal direction and the vertical eye movement requirement is a maximum of 1 degree. Possibly, for eyes with better VA and a steeper

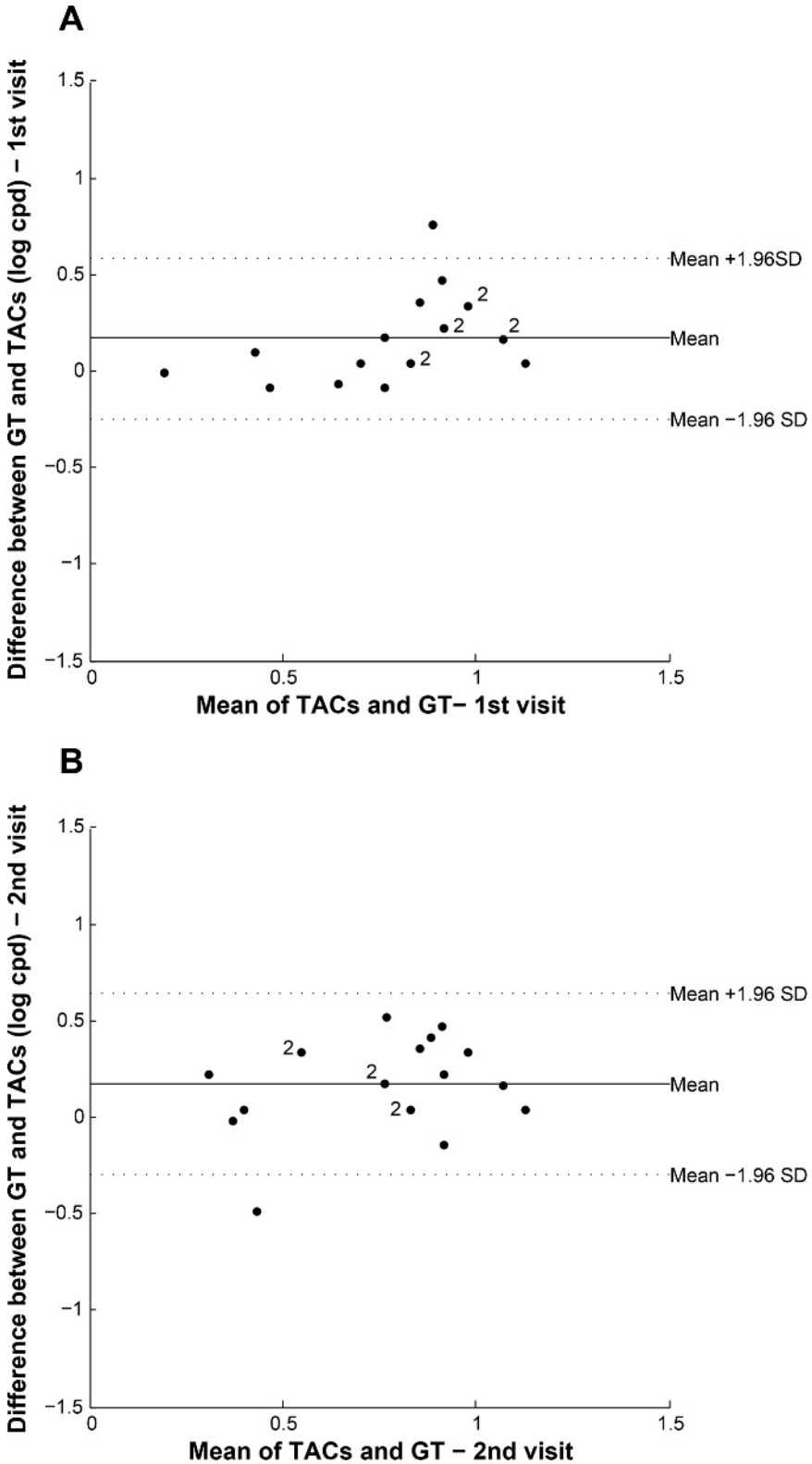

FIGURE 5.

Scatterplots of VA obtained with the GT against TACs used with horizontal gratings for infant participants. (A) First visit. (B) Second visit. The difference between acuities is plotted against the mean of the acuities. The digits indicate the number of data points where there is more than one data point in a given location. The dotted lines indicate the 95% limits of agreement.

decrease of VA extrafoveally, there is more chance that the infant will detect the grating. Interestingly, the adult data indicate no systematic bias between the GT and the TACs.

The 89% to less than or equal to 1 octave repeatability of the GT over two visits was similar to acuity card studies in the infant literature. McDonald et al.[16] found an intraobserver repeatability

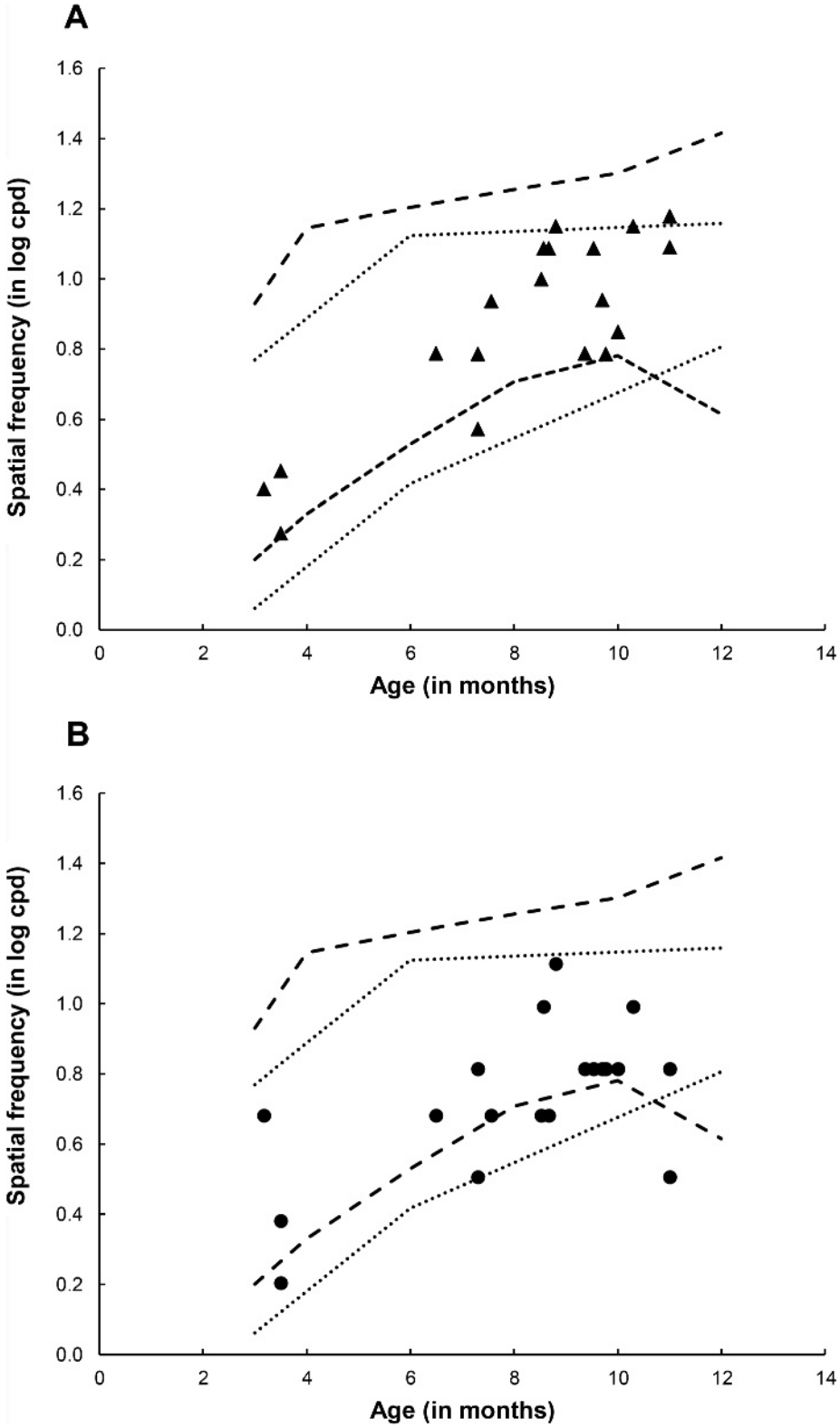

FIGURE 6.

Scatterplots of average grating acuities plotted against age. (A) Gaze tracking VA. (B) Teller Acuity Cards. In both A and B, the dashed lines are the upper and lower 95% confidence limits for TACs adapted from Salomao and Ventura[22] and the dotted lines are the upper and lower 95% range based on data from Courage and Adams.[5]

of 87.5% within 1 octave between visits, similar to the results in this study for both the GT and TACs.

Limitations

For the adult data, the observer was not blind to the spatial frequency of the gratings and therefore some bias may have occurred in repeated measures. This is unlikely because the point in the staircase where adults stopped finding and fixating the gratings was very clear.

One of the limitations of the infant study was the possible change in test distance, if the baby moved closer or farther from the gratings. The error resulting from this change is negligible for larger test distances of the GT. For example, a 5-cm change in testing distance would cause a 0.08-log unit change in TACs at 55 cm, whereas the same change in distance would cause a 0.05-log unit change in the GT at the 120-cm working distance. Both of these are less than one step in the current study and less than one step on most logMAR charts. A second limitation is that the testing distance between TACs and the GT was not the same. This was because the pixels of the screen limit resolution and thus the GT video cameras were calibrated for a distance of 70 cm or greater. Thus, it is possible that a greater error of focus may have been present for one distance compared with the other. Another limitation is that we did not have infants between 3.5 and 6.5 months in this initial proof-of-principle study. Last, we did not measure monocular VA, which is important clinically. Although not demonstrated in this study, it should be noted that the GT can also track a single eye, so that there is the potential for measuring monocular acuity.

CONCLUSIONS

We have shown that the GT using naïve eye movement responses gives a valid measure of grating acuity in adults, indicating the potential for measuring VA in infants. We then demonstrated good testability, repeatability, and validity of the GT in infants, thus demonstrating the potential of the GT as an objective measure of VA in infants. Because the decision of the observer in the present study was based only on the data provided by the gaze tracker, we conclude that the development of fully automated algorithms is feasible, which would provide a more objective VA measurement. Such a system may be useful in screening for patients who have difficulty responding and could be modified to test visual functions other than VA.

ACKNOWLEDGMENTS

This work was funded by Collaborative Health Research Projects (CHRP) Canada.

Portions of the research discussed in this article were presented at the Association for Research in Vision and Ophthalmology meeting, Seattle, May 2013; the American Academy of Optometry annual meeting, Phoenix, October 2012; and the XIV Biennial Meeting of the Child Vision Research Society, Waterloo, Canada, 2013.

Received August 16, 2014; accepted March 3, 2015.

REFERENCES

1. Banks MS, Salapatek P. Acuity and contrast sensitivity in 1-, 2-, and 3-month-old human infants. Invest Ophthalmol Vis Sci 1978; 17:361–5.
2. 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;28:779–89.
3. Gwiazda J, Wolfe JM, Brill S, Mohindra I, Held R. Quick assessment of preferential looking acuity in infants. Am J Optom Physiol Optic 1980;57:420–7.
4. Dobson V. Visual acuity testing in infants: from laboratory to clinic. In: Simons K, ed. Early Visual Development, Normal and Abnormal. New York, NY: Oxford University Press; 1993:318–34.
5. 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;67:713–8.
6. McDonald M, Ankrum C, Preston K, Sebris SL, Dobson V. Monocular and binocular acuity estimation in 18- to 36-month-olds: acuity card results. Am J Optom Physiol Opt 1986;63:181–6.
7. Quinn GE, Berlin JA, James M. The Teller acuity card procedure. Three testers in a clinical setting. Ophthalmol 1993;100:488–94.
8. 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: 671–85.
9. Sturm V, Cassel D, Eizenman M. Objective estimation of visual acuity with preferential looking. Invest Ophthalmol Vis Sci 2011; 52:708–13.
10. Guestrin ED, Eizenman E. General theory of remote gaze estimation using the pupil center and corneal reflections. IEEE Trans Biomed Eng 2006;53:1124–33.
11. Guestrin ED, Eizenman M. Remote point-of-gaze estimation requiring a single-point calibration for applications with infants. In: ETRA '08 Proceedings of the 2008 Symposium on Eye Tracking Research & Applications. New York, NY: ACM; 2008:267–74.
12. Model D, Eizenman M. User-calibration-free remote eye-gaze tracking system with extended tracking range. In: Proceedings of the 24th Canadian Conference on Electrical and Computer Engineering (CCECE). Piscataway, NJ: IEEE; 2011:1268–71.
13. Model D, Eizenman M. An automated Hirschberg test for infants. IEEE Trans Biomed Eng 2011;58:103–9.
14. Teller DY, Dobson V, Mayer DL. Teller Acuity Cards II Handbook. Chicago, IL: Stereo Optical Company Inc; 2005.
15. Hathibelagal A. Objective assessment of visual acuity in infants. MS thesis. Waterloo, ON, Canada: University of Waterloo; 2013. Available at: https://uwspace.uwaterloo.ca/handle/10012/7681. Accessed March 28, 2015.
16. 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;26:1158–62.
17. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;327: 307–10.
18. Lewis TL, Reed MJ, Maurer D, Wyngaarden PA, Brent HP. An evaluation of acuity cards procedure. Clin Vis Sci 1993;8:591–602.
19. Gwiazda J, Thorn F, Bauer J, Held R. Emmetropization and the progression of manifest refraction in children followed from infancy to puberty. Clin Vis Sci 1993;8:337–44.
20. Gwiazda J, Grice K, Held R, McLellan J, Thorn F. Astigmatism and the development of myopia in children. Vision Res 2000;40: 1019–26.

21. Clifford-Donaldson CE, Haynes BM, Dobson V. Teller Acuity Card norms with and without use of a testing stage. J AAPOS 2006; 10:547–51.
22. Salomao SR, Ventura DF. Large sample population age norms for visual acuities obtained with Vistech-Teller Acuity Cards. Invest Ophthalmol Vis Sci 1995;36:657–70.
23. Lovie-Kitchin JE, Brown B. Repeatability and intercorrelations of standard vision tests as a function of age. Optom Vis Sci 2000; 77:412–20.
24. Plainis S, Kontadakis G, Feloni E, Giannakopoulou T, Tsilimbaris MK, Pallikaris IG, Moschandreas J. Comparison of visual acuity charts in young adults and patients with diabetic retinopathy. Optom Vis Sci 2013;90:174–8.
25. Mohn G, van Hof-van Duin J, Fetter WP, de Groot L, Hage M. Acuity assessment of non-verbal infants and children: clinical experience with the acuity card procedure. Dev Med Child Neurol 1988;30:232–44.
26. Preston KL, McDonald M, Sebris SL, Dobson V, Teller DY. Validation of the acuity card procedure for assessment of infants with ocular disorders. Ophthalmology 1987;94:644–53.

Amithavikram R. Hathibelagal
8/308 Kelvin Grove Rd
Kelvin Grove, Brisbane
Queensland 4059
Australia
e-mail: amithrugvedi@gmail.com