Article

Testing Pediatric Acuity With an iPad: Validation of "Peekaboo Vision" in Malawi and the UK

Iain Livingstone[1,2], Laura Butler[3], Esther Misanjo[4,5], Alan Lok[3,6], Duncan Middleton[7], Janice Waterson Wilson[8], Silvija Delfin[9], Petros Kayange[4,5], and Ruth Hamilton[8,10]

[1] Falkirk Community Hospital, NHS Forth Valley, Falkirk, UK
[2] Glasgow Centre for Ophthalmic Research, NHS Greater Glasgow and Clyde, Glasgow, UK
[3] Tennent Institute of Ophthalmology, Gartnavel General Hospital, NHS Greater Glasgow and Clyde, Glasgow, UK
[4] Lions First Sight Eye Unit, Queen Elizabeth University Hospital, Blantyre, Malawi
[5] Ophthalmology Department, College of Medicine, University of Malawi, Blantyre, Malawi
[6] University Hospitals Coventry and Warwickshire NHS Trust, Coventry, UK
[7] Medical Devices Unit, West Glasgow Ambulatory Care Hospital, NHS Greater Glasgow and Clyde, Glasgow, UK
[8] Royal Hospital for Children, NHS Greater Glasgow and Clyde, Glasgow, UK
[9] Inverclyde Royal Hospital, NHS Greater Glasgow and Clyde, Greenock, UK
[10] College of Medical, Veterinary & Life Sciences, University of Glasgow, Glasgow, UK

Purpose: To evaluate two builds of the digital grating acuity test, "Peekaboo Vision" (PV), in young (6–60 months) populations in two hospital settings (Malawi and United Kingdom).

Methods: Study 1 evaluated PV in Blantyre, Malawi ($N = 58$, mean age 33 months); study 2 evaluated an updated build in Glasgow, United Kingdom ($N = 60$, mean age 44 months). Acuities were tested-retested with PV and Keeler Acuity Cards for Infants (KACI). Bland-Altman techniques were used to compare results and repeatability. Child engagement was compared between groups. Study 2 included test-time comparison.

Results: Study 1 (Malawi): The mean difference between PV and KACI was 0.02 logMAR with 95% limits of agreement (LoA) of 0.33 to 0.37 LogMAR. On test-retest, PV demonstrated 95% LoA of −0.283 to 0.198 logMAR with coefficient of repeatability (CR) 0.27. KACI demonstrated 95% LoA of −0.427 to 0.323 logMAR, and larger CR was 0.37. PV evidenced higher engagement scores than KACI ($P = 0.0005$). Study 2 (UK): The mean difference between PV and KACI was 0.01 logMAR; 95% LoA was −0.413 to 0.437 logMAR. Again, on test-retest, PV had narrower LoA (−0.344 to 0.320 logMAR) and lower CR (0.32) versus KACI, with LoA −0.432 to 0.407 logMAR, CR 0.42. The two tests did not differ in engagement score ($P = 0.5$). Test time was ~1 minute shorter for PV (185 vs. 251 s, $P = 0.0021$).

Conclusions: PV gives comparable results to KACI in two pediatric populations in two settings, with benefits in repeatability indices and test duration.

Translational Relevance: Leveraging tablet technology extends reliable infant acuity testing to bedside, home, and rural settings, including areas where traditional equipment cannot be financed.

Introduction

Almost half of childhood blindness is preventable or treatable.[1] Early treatment supports normal development of the visual brain, avoiding amblyopia. Sixty-three percent of childhood visual impairment worldwide is due to refractive error, and treatment during the phase of rapid development[2,3] helps avoid permanent impairment. The first 5 years of life are therefore an important window for effective screening.

Pediatric acuity testing traditionally uses forced-choice, preferential-looking gratings printed on card.[4]

The advent of tablet computers creates the opportunity to test acuity digitally. Tablets cost less than card-based acuity tests and can emulate vision tests as games. Automated stair-casing and reporting require fewer and less-specialized testers and could extend the reach of visual screening programs in both developed nations and in economies with limited access to health care.

The present study evaluates a digital, tablet-based forced-choice acuity test, Peekaboo Vision, designed to function as a preferential looking or touchscreen game where the child can interact directly with the screen. We evaluated the technology in two pediatric populations, extending work in a blurred adult cohort.[5] A prototype build was first assessed in a cohort in Blantyre, Malawi (study 1). This pilot study allowed the concept of iPad-based (Apple, Cupertino, CA) digital gratings to be tested against a known reference standard and was instructive regarding the subsequent formal software development. Study 1 also served as a test of methodology, guiding the design of study 2, where the formal build with a similar paradigm was evaluated in a United Kingdom cohort.

Differences between the heterogeneous test populations, together with multiple intrinsic divergences in application builds (target design, stair-casing, testable range, and distance) make direct comparison between studies 1 and 2 somewhat spurious. Data are hence presented separately for each study.

Methods and Materials

The research followed the tenets of the Declaration of Helsinki. Informed consent was obtained from the subjects' parents/guardian after explanation of the nature and possible consequences of the study.

Study 1

Patients

The Malawi College of Medicine Research and Ethics Committee granted ethical approval. Fifty-eight consecutive, unselected children, aged 6–60 months, presenting to the Lions Sight First Eye Unit in Blantyre were recruited, many of whom exhibited visual problems.

Setting

Testing was performed in clinic rooms by an ophthalmologist (LB or EM) experienced in testing pediatric vision. Testers were masked to any docu-mented clinical information other than the age and date of birth of the child. This included diagnosis, past medical history, and any previous visual measurements such as acuity, visual field, refraction, or orthoptic problems. The same tester performed all tests on each child. Intertester variability was not investigated. All testing was performed in well-lit clinical areas by clinical staff with experience in testing children's vision. As the focus of the validation was to evaluate performance in a real-life environment against the present reference standard, no luminance measurements were taken regarding ambient light levels or luminance levels from the Keeler cards. Previous research has investigated the luminance conformity of the iPad as an instrument for measuring acuity.[6]

All children were tested with both Peekaboo Vision and Keeler cards in random order. If the *day* of month within the date of birth of the child had an even number, Keeler was performed first. The both-eyes-open (BEO) condition was undertaken first. If the child had an even *month* of birth, then the right eye was tested next; otherwise, it was left first. Typical occlusive glasses designed for pediatric acuity testing[7] were used to cover the fellow eye in the right eye (RE) and left eye (LE) tests. A handheld monocular occluder was also used for younger children who did not tolerate occlusive spectacles.

Peekaboo Vision Build 1 (PVb1)

Study 1 used PVb1, a prototype previously described.[5] Graphics used Adobe Photoshop Creative Cloud (Adobe Inc., San Jose, CA), and the application was developed in HTML5 to screen specification of third-generation iPad with "retina display" (resolution 264 pixels per inch). Vertical grating targets were employed, being the most robustly described clinical standard for infants.[8] To engage and hold attention, grating targets comprised simple smiley faces (Fig. 1A) against an isoluminant background generated by an alternating black/white checkerboard at the maximum resolution of the display, which appeared as uniform gray. The detail of the eyes/mouth to comprise the smiley face elements were composed of the same isoluminant checkerboard pattern used to generate the background, such that they would be visible only if the grating could be delineated from the background. Test distance was 25 cm, with an acuity range of 0.1 to 2.2 logMAR (Table 1). Screen brightness was manually set to 50%. Measured acuity thresholds were expressed in the logarithm of the minimum angle of resolution.
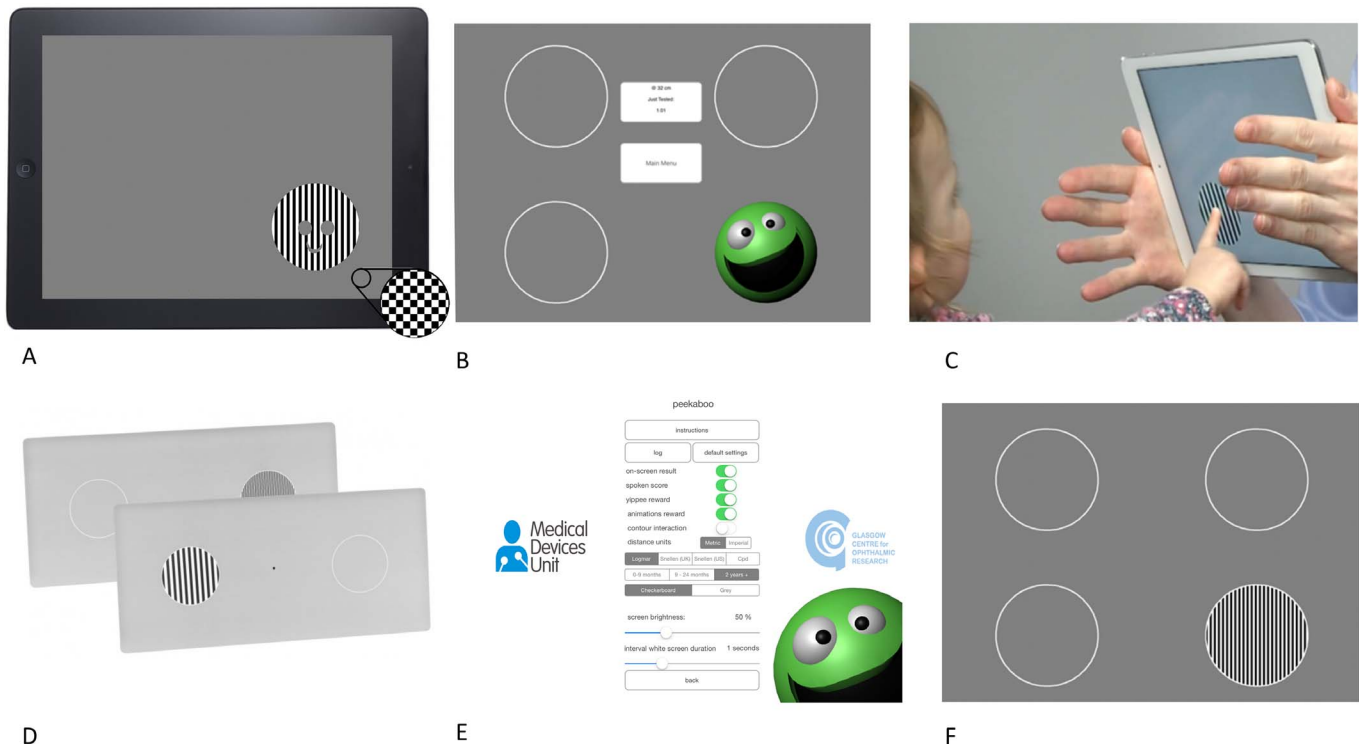
Figure 1. (A) PVb1. (B) Reward graphic and on-screen result. (C) Demonstration of touch screen functionality. (D) The Keeler Acuity Test for Infants. (E) Peekaboo Vision build for iOS (PVi) option configuration screen. (F) PVi example test screen.

The tablet was held in landscape orientation facing the child so that the tester was masked to target location. Targets appeared pseudorandomly in one quarter of the screen following a 0.3 logMAR down, 0.1 logMAR up the staircase (Fig. 2A). Infants' eye movements were used to infer the perceived location of the grating, with the tester tapping the corresponding screen quarter. Older children touched or pointed to "the smiley face" grating target (Fig. 1A). Correct results produced a sound/animation reward involving a smiley face (Fig. 1B). The tester held the tablet with the fingers positioned such that the test screen was not obscured (Fig. 1C). In the absence of a correct response, a lower-frequency target was presented.

Study Protocol

Keeler Acuity Cards for Infants (KACI) were selected as the reference standard in this validation study, as it is the chosen test in our tertiary referral center in Scotland and is used widely in UK practice. It is also the reference standard in other evaluations of digital grating paradigms.[9] While KACI does not have as extensive normative data as others, specifically Teller Acuity Cards,[8,10] despite differences in design between both tests, Neu et al.[11] reported KACI and Teller cards compared favorably on monocular

testing in a slightly older, but similar age group to the present study (1–6 years old, $N = 95$), with no significant differences in acuity scores.

An 18-card set was used, covering −0.1 to 2.0 logMAR (Fig. 1D, range detailed in Table 1), including a blank card with no grating. The 1.5-logMAR card was used first, moving up/down the staircase depending on responses, as per instructions for use. For younger children, looking responses were judged by the tester.

Table 3 details the sizing of the grating elements for KACI and Peekaboo Vision.

All 58 children were tested with both Peekaboo Vision and Keeler cards. The BEO condition was undertaken first, followed by RE/LE conditions, totaling six acuity tests. After a 15-minute interval, this process was repeated to assess test-retest repeatability. Each of the 58 children therefore underwent 12 acuity tests in total. Target distance was maintained using premeasured marks on the tester's arm. An engagement score (ES) of 0, 1, or 2 was awarded for every test result: 0 = no meaningful results; 1 = some meaningful results but loss of interest during test; and 2 = engagement to convincing threshold or finest grating.

Table 1. Spatial Frequencies and Equivalent Acuity (Logmar) of Gratings Available With the Keeler Acuity Cards for Infants Test Plus Additional Cards and With Each of the Builds of the Peekaboo Vision Test

Target Spatial Frequency, cpd	Acuity, logMAR Equivalent	Keeler Cards, 38 cm, Additional Set*	PVb1 at 25 cm	PVi at: 25 cm	30 cm	40 cm	50 cm
51.0	−0.2						✓
35.4	−0.1	*				✓	
29.0	0.0				✓		
25.5	0.1	*	✓	✓			✓
21.6	0.2	*				✓	
14.5	0.3	*			✓		✓
12.5	0.4	✓	✓	✓		✓	✓
9.6	0.5	*			✓	✓	✓
7.7	0.6	*	✓	✓	✓	✓	✓
6.5	0.7	✓	✓	✓	✓	✓	✓
4.8	0.8		✓	✓	✓	✓	✓
3.8	0.9	*	✓	✓	✓	✓	✓
2.9	1.0	✓	✓	✓	✓	✓	✓
2.4	1.1		✓	✓	✓	✓	✓
2.1	1.2	*	✓	✓	✓	✓	✓
1.4	1.3	✓	✓	✓	✓	✓	✓
1.2	1.4		✓	✓	✓	✓	✓
1.0	1.5	*	✓	✓	✓	✓	✓
0.7	1.6	✓	✓	✓	✓	✓	✓
0.6	1.7		✓	✓	✓	✓	✓
0.5	1.8		✓	✓	✓	✓	✓
0.4	1.9	✓	✓	✓	✓	✓	✓
0.3	2.0	*	✓	✓	✓	✓	✓
0.2	2.2	✓	✓	✓	✓	✓	✓

cpd, cycles per degree.

Study 2

Methods for study 2 matched those for study 1, except for the aspects detailed below. Time to test was recorded for each test and viewing condition (BEO, RE, LE).

Patients

The West of Scotland Research Ethics Committee granted ethical approval. Sixty unselected children, aged 2–60 months, were recruited from the Royal Hospital for Sick Children eye clinic (Glasgow, UK), including those who exhibited visual problems as well as siblings free from visual problems.

Setting

Testing was performed in clinic rooms by an ophthalmologist (IL, SD) or senior orthoptist (AL, JWW).

Peekaboo Vision Formal Build for iOS (PVi)

A formal build used Swift 2 language for iOS versions 8.1 and above (Apple, Cupertino, CA), scaled for the iPad3 platform. A video demonstration of this build is available in the supplementary material (Supplementary Video S1). The options screen is shown in Figure 1E. Changes from build 1 were informed by in-house testing and study 1, comprising the following:

- Randomization (rather than pseudo-randomization) of grating location;
- Automatic (rather than manual) overriding of brightness settings;
- Reentry onto the rapid staircase if three consecutive levels are correct after an initial error (Fig. 2B) to compensate for accidentally incorrect responses, sometimes caused by the tester inadvertently touching the screen;

A

Test Start (Default: 1.3 logmar)

correct response → incorrect response

descends 1 octave steps *down* staircase (*rapid staircase*). **Continues until incorrect response *or* finest grating**

incorrect response → ascends in 0.1 logmar steps *up* staircase (*slow staircase*)

correct response / incorrect response

If 2 out of 2 presentations correct, threshold reached

Continues until threshold: 2 out of 2 correct or "no level passed"

B

Test Start (Default: 1.3 logmar)

correct response → incorrect response

descends 1 octave steps *down* staircase (*rapid staircase*). **Continues until incorrect response *or* finest grating**

incorrect response → ascends in 0.1 logmar steps *up* staircase (*slow staircase*)

correct response / incorrect response

descends in 0.1 logmar steps *down* staircase (*slow staircase*)

Continues until threshold: 2 out of 3 correct or "no level passed"

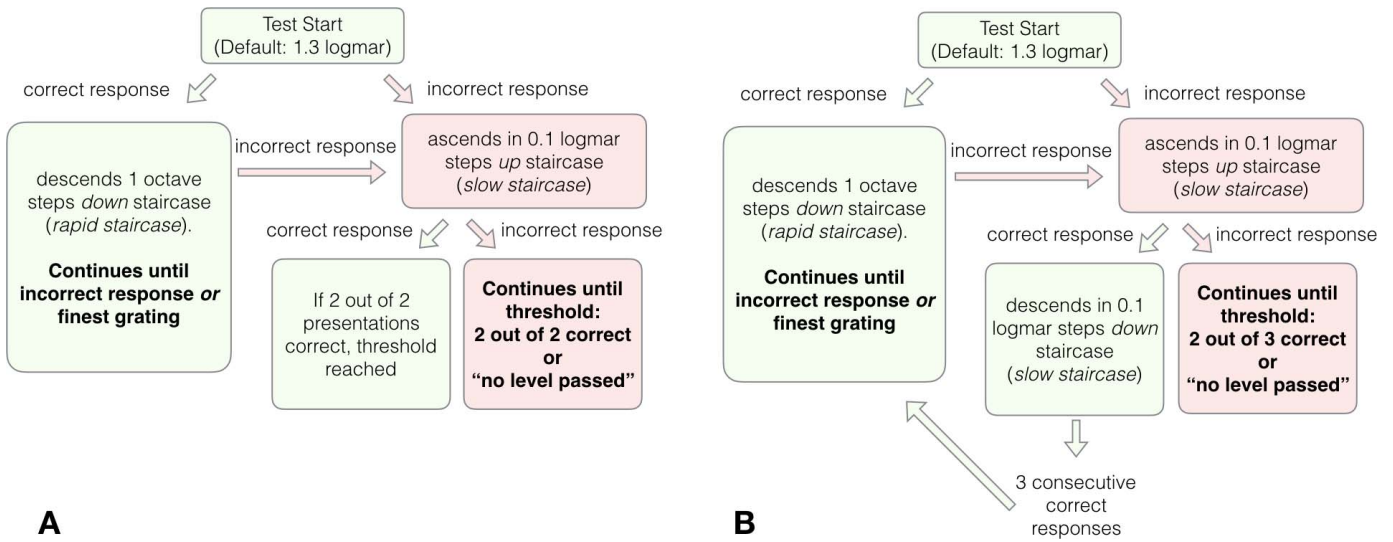3 consecutive correct responses

Figure 2. (A) Stair-casing paradigm used in PVb1. Correct responses occur when the child taps or points to the correct corner of the tablet or when the tester observes purposeful eye gaze and taps the corresponding corner of the tablet. Incorrect responses occur when child taps or points to the wrong corner, when the tester observes purposeful eye gaze and taps the incorrect corresponding corner, or if there are no meaningfully directed eye movements. (B) Stair-casing paradigm used in PVi. The stair-casing change from the PVb1 version comprises automatic reentry into the *rapid staircase* following three consecutive correct responses.

Table 2. Physical Properties/Sizing Details for KACI and Peekaboo Vision

	KACI		Peekaboo Vision		
		Visual Angle		Visual Angle Subtended, arcmin, at:	
	Physical Size, mm	Subtended at 38 cm, arcmin	Physical Size, mm	25 cm[a]	50 cm (PVi)[a,b]
Grating patch diameter	103	931	48	660	330
Grating patch center to center	155	1402	H: 100	H: 1375	H: 688
			V: 76	V: 1045	V: 523
			D: 126	D: 1732	D: 866

H, horizontal; V, vertical; D, diagonal.
[a] Indicates the distance used in both PVb1 and PVi builds.
[b] Extended distance used at limit of dynamic range in PVi, study 2.

Table 3. Number of Tests (Comprising BEO, RE, and LE) With Each Engagement Score (0, 1, or 2) for the Two Acuity Tests

		Engagement Score			P Value of Difference in Proportions of Score = 2
		0	1	2	
Study 1 (N = 58 subjects)	Keeler cards (N = 174 measures)	16	79	79	0.0005
	PVb1 (N = 174 measures	11	63	100	
Study 2 (N = 60 subjects)	Keeler cards (N = 158 measures)	7	28	123	0.5
	PVi (N = 158 measures	8	32	118	

Figures are for first test, not retest. In the minority of instances where children had an abject loss of engagement or became upset, no further testing was performed.

- Removal of the smiley face details that created subtly visible edge artefacts within the grating (Fig. 1A);
- Addition of a ring around the grating target (Fig. 1F), similar to Keeler cards, to limit visual cues where high-frequency gratings interact with the background;
- Configurable test distance, adjustable at centimeter increments within the range 25 to 50 cm (rather than fixed at 25 cm), to bypass screen resolution limitations, creating −0.2, −0.1, 0.0, 0.2, and 0.3 logMAR levels (Table 1);
- Addition of a beep with each new target presentation to aid tester's recognition of synchronous looking responses coincident with appearance of grating; and
- Addition of a feature to re-present the same target at a new random location in cases of equivocal responses by shaking the device (akin to twirling a Keeler card).

Statistics (Studies 1 and 2)

Engagement scores for the two test formats within each study (Keeler cards versus PVb1 or PVi) were compared using McNemar's test for correlated proportions. Engagement scores (BEO versus monocular testing and also test 1 versus retest) were compared using Mann-Whitney U tests. For test 1 versus retest comparison, BEO/RE/LE conditions were grouped together into test and retest groups.

Acuity scores were summarized using median values and compared between tests using limits of agreement and Bland-Altman plots. Test-retest repeatability was described using limits of agreement and CR (twice the standard deviation of the differences).

Test time was compared between PVi and KACI in study 2 using a paired t-test.

Results

Study 1

Engagement Scores

Peekaboo Vision had higher engagement scores than did Keeler cards over all three viewing conditions (BEO, RE, LE) (Table 3). An engagement score of 2 was achieved on significantly more occasions for Peekaboo Vision than for Keeler cards (100/174 [57%] versus 79/174 [45%]; McNemar's test for correlated proportions, $P = 0.0005$).

Engagement score with respect to age was analyzed for BEO results for both tests. Thirty-six children attained an engagement score of 2 with Peekaboo Vision; their median age was 37 months. Thirty-one children attained an engagement score of 2 with Keeler cards; their median age was 38 months. Engagement scores of 1 were more frequent in younger children, with median ages of 28 months ($N = 23$) and 25.5 months ($N = 18$) for Keeler cards and Peekaboo Vision, respectively. Engagement scores of 0 were infrequent, but in study 1 occurred in the same older children ($N = 4$), with a median age of 36.5, using both Keeler cards and Peekaboo Vision.

There was evidence that engagement dropped slightly for Keeler card monocular testing following binocular testing: the average engagement score dropped from median 2 (mean 1.5) for binocular testing ($N = 58$) to median 1 (mean 1.3) for monocular testing ($N = 58$, Mann-Whitney U test, W [$W =$ Mann-Whitney test statistic: the sum of the ranks of the first value] $= 3665$, adjusted $P = 0.10$). With Peekaboo Vision testing, engagement seemed to be marginally better maintained: average engagement score for binocular testing ($N = 58$) dropped from median 2 (mean 1.6) to median 1.5 (mean 1.5) for monocular testing ($N = 58$, Mann-Whitney U test, $W = 3565$, adjusted $P = 0.3$).

For Keeler cards, engagement score did not significantly drop between test and retest, with a median score of 1 for both groups, with a modest decrease in the mean from 1.4 for test 1 versus 1.2 for retest ($N = 174$, Mann-Whitney U test, $W = 31,977$, adjusted $P = 0.06$).

For PVb1, ES did not significantly change between test and retest, but evidenced a maintained median ES of 2 for across test and retest, with the mean decreasing from 1.5 to 1.4 on retest ($N = 174$, Mann-Whitney U test, $W = 31,192$, adjusted $P = 0.32$).

Acuity Thresholds

For all viewing conditions (BEO, RE, LE), only children attaining an engagement score of 2 for both Peekaboo Vision and Keeler cards were included in comparative analysis. Seventy-two tests from 31 children had engagement scores of 2 for both Peekaboo Vision and for Keeler cards. Median Peekaboo Vision acuity was 0.5 logMAR (range, 0.1–1.9); median Keeler cards acuity was 0.4 logMAR (range, 0.1–1.6). Agreement between the two tests was good, with a median absolute difference of 0.1 logMAR, mean difference of 0.02, 95% LoA −0.33
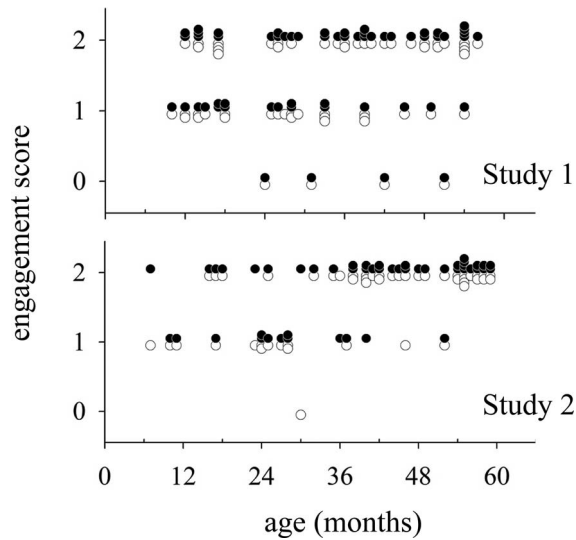
Figure 3. Scatterplots showing distribution of engagement scores with age of subjects for study 1 (*upper panel*) and study 2 (*lower panel*). Data are for first tests, BEO viewing condition. *Open circles*: Keeler cards. *Closed circles*: Peekaboo Vision.

to 0.37 LogMAR. The acuity difference between the two tests was ≤0.4 logMAR in 95% of tests and showed no tendency to vary with acuity (Fig. 4A).

Test-Retest

To assess test-retest repeatability, pairs of tests with an engagement score of 2 at first test and again at retesting some 15 minutes later, and for all viewing conditions (RE, LE, BEO), were compared. Eighty-five pairs of Peekaboo Vision acuities and 58 pairs of Keeler card acuities were included (Fig. 4B). Both tests showed good repeatability, with median differences between test and retest being zero. Mean test-retest differences of −0.042 logMAR and −0.052 logMAR were found for Peekaboo Vision and for Keeler cards, respectively. LoA were narrower, and CR were lower (better) for Peekaboo Vision (LoA −0.283 to 0.198 logMAR, CR 0.27) when compared to Keeler cards (LoA −0.427 to 0.323 logMAR, CR 0.37).

Repeatability was similar for Keeler binocular testing ($N = 20$, LoA −0.359 to 0.399 logMAR, CR 0.39) and monocular testing ($N = 38$, LoA −0.308 to 0.435 logMAR, CR 0.38). Repeatability was slightly poorer for Peekaboo binocular testing ($N = 31$, LoA −0.270 to 0.315 logMAR, CR 0.30) than for monocular testing ($N = 54$, LoA −0.148 to 0.244 logMAR, CR 0.20).

Test-retest acuity differences were compared for Keeler cards ($N = 20$) and for Peekaboo vision ($N = 31$) for BEO viewing conditions and engagement

scores of 2: no significant differences were found (mean test-retest differences 0.020 versus 0.023 logMAR, 95% confidence interval (CI) of difference −0.09 to 0.09, $P = 0.95$).

Study 2

Engagement Scores

Peekaboo Vision and Keeler cards had very similar engagement scores (Table 3) based on the first test result for all three viewing conditions (BEO, RE, LE). An engagement score of 2 was achieved for slightly fewer Peekaboo Vision tests than for Keeler card tests (118/158 [75%] versus 123/158 [77%]); McNemar's test for correlated proportions, $P = 0.5$). Engagement scores for BEO results were reviewed with respect to age as for study 1. Forty-seven children attained an engagement score of 2 with Peekaboo Vision; their median age was 54 months. Forty-five children attained an engagement score of 2 with Keeler cards; their median age was also 54 months. As for study 1, engagement scores of 1 were more frequent in younger children, with median ages of 27 months for both Peekaboo Vision ($N = 13$) and Keeler cards ($N = 14$). Engagement score of zero occurred only once, for a 17-month-old) child using Keeler cards.

There was no convincing evidence that engagement dropped for monocular testing following binocular testing with Keeler cards: average engagement score for binocular testing ($N = 49$) was median 2 (mean 1.8) and was median 2 (mean 1.7) for monocular testing ($N = 49$, Mann-Whitney U test, $W = 2505$, adjusted $P = 0.4$). For Peekaboo Vision testing, engagement dropped slightly: average engagement score for binocular testing ($N = 49$) was median 2 (mean 1.9) and was median 2 (mean 1.6) for monocular testing ($N = 49$, Mann-Whitney U test, $W = 2668$, adjusted $P = 0.02$).

Regarding change in engagement on test-retest, for Keeler cards the median ES was maintained at 2, with the mean ES dropping modestly from 1.9 to 1.7 ($N = 119$, Mann-Whitney U test, $W = 14,859$, adjusted $P = 0.07$).

Similarly, PVi did not show a significant change in ES on retest, with median ES of 2 in both groups and the mean decreasing from 1.9 to 1.8 on retest ($N = 119$, $W = 14,763$, adjusted $P = 0.10$).

Time to Test

The time-to-test data reflects the first test performed to limit any bias from learning or loss of interest from prolonged testing. Only those tests attaining engagement score of 2 (convincingly reached
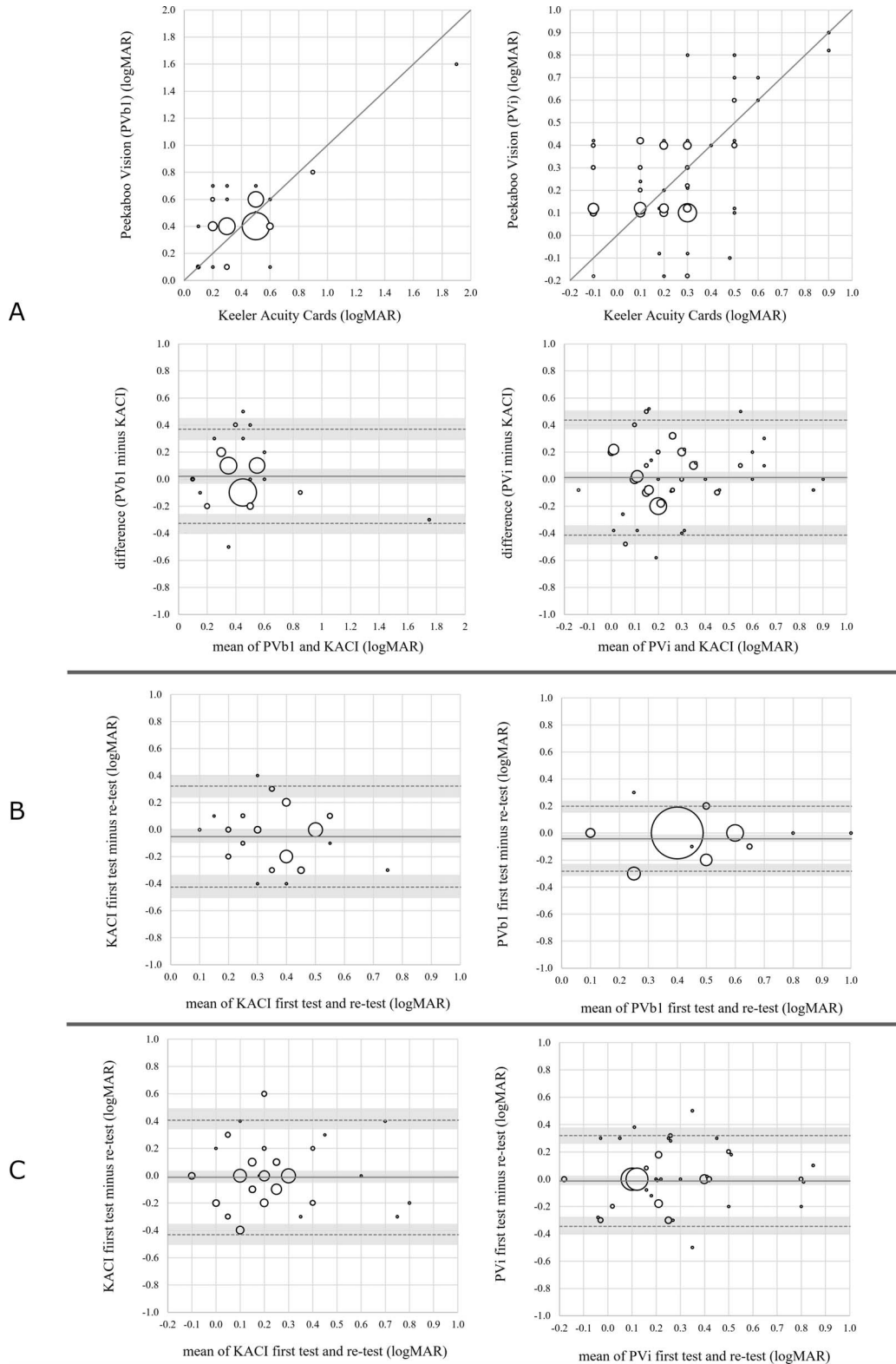
Figure 4. (A) Acuities recorded in study 1 (Malawi, PVb1; *left panels*) and study 2 (United Kingdom, PVi; *right panels*). *Upper panels* plot Peekaboo Vision acuities versus Keeler cards acuities for all acuities with an engagement score of 2. *Lower panels* show Bland-Altman plots of agreement between the two different tests. *Circles* represent data points and are scaled to represent the number of instances the
→

threshold without losing interest) with BEO were included. Mean time to test was just over a minute shorter for Peekaboo Vision than for Keeler cards ($N = 33$, 185 vs. 251 s; paired *t*-test, $P = 0.002$).

Acuity Thresholds

For all viewing conditions (BEO, RE, LE), only those tests attaining an engagement score of 2 for both Peekaboo Vision and for Keeler cards were included in analysis. One hundred ten tests from 46 infants and children had engagement scores of 2 for both Peekaboo Vision and Keeler cards. Median Peekaboo Vision acuity was 0.2 logMAR (range, −0.18 to 0.90); median Keeler card acuity was 0.3 logMAR (range, 0.10–0.90). Agreement between the two tests was good, with a median absolute difference of 0.18 logMAR (mean difference: 0.01, 95% LoA −0.413 to 0.437 logMAR). As would be expected in this population, the most frequently encountered acuities were in the normal range, between 0.0 and 0.4 logMAR, with no obvious bias across the range of acuities encountered (−0.18 to 0.9; Fig. 4A, bottom right panel).

Test-Retest

As for study 1, all pairs of tests with an engagement score of 2 at first test and again at retesting and for all viewing conditions (BEO, RE, LE) were assessed. Ninety-one pairs of Peekaboo Vision acuities and 90 pairs of Keeler card acuities (Fig. 4C) were compared. Both tests showed good repeatability, with median differences between test and retest of zero and mean differences of −0.012 logMAR for both Peekaboo Vision and Keeler cards. As in study 1, LoA were narrower and CR were lower (better) for Peekaboo Vision (LoA −0.344 to 0.320 logMAR, CR 0.32) than for Keeler cards (LoA −0.432 to 0.407 logMAR, CR 0.42).

Repeatability was similar for KACI binocular testing ($N = 36$, LoA −0.429 to 0.429 logMAR, CR 0.44) and monocular testing ($N = 54$, LoA −0.411 to 0.433 logMAR, CR 0.43). Repeatability was also similar for PVi binocular testing ($N = 33$, LoA −0.262 to 0.383 logMAR, CR 0.33) and monocular testing ($N = 58$, LoA −0.338 to 0.298 logMAR, CR 0.32).

Test-retest acuity differences were compared for KACI ($N = 36$) and for PVi ($N = 33$) for BEO viewing conditions and engagement scores of 2: no significant differences were found (mean test-retest differences 0.00 vs. 0.06 logMAR, 95% CI of difference −0.15– 0.03, $P = 0.2$).

Discussion

In both studies, the mean difference in acuities measured by the card-based test and by the digital test is modest, being 0.02 logMAR (95% CI for mean difference: −0.02 to 0.06) for PVb1 (study 1, Malawi) and 0.01 logMAR (95% CI for mean difference: −0.03 to 0.05) for PVi build (study 2, United Kingdom). When comparing the index with the reference standard, the upper and lower limits of agreement (the interval of two standard deviations of the measurement differences either side of the mean difference) exceeded an octave step, but they were within 5 logMAR lines (LoA 0.33–0.37 logMAR for study 1; 95% LoA −0.413 to 0.437 logMAR for study 2). These limits of agreement are similar to those observed when KACI is compared with itself on retest in both studies (study 1: LoA −0.427 to 0.323 logMAR; study 2: LoA −0.432 to 0.407 logMAR).

Furthermore, with narrower limits of agreement on test-retest when compared with KACI in both studies, the findings support the use of high-resolution tablet-based technology, such as iPads, as a credible addition to the armamentarium available to clinicians in the assessment of vision in young children. However, as discussed below, there are limitations within these studies, and it cannot be unequivocally concluded that that these digital tests represent a replacement of the reference standard.

The number of forced-choice alternatives represents a major difference between the index tests and the two-target Keeler Acuity Card. With PVb1, the staircase continued until two out of two presentations were correct. This rigid stair-casing did not allow progression when an error was made, and only poorer levels were tested thereafter. This method of stair-casing reflects the recommended testing strategy for Keeler cards, as per instructions for use. Assuming a one in four (0.25) probability of the correct target

being selected by pure chance at one presentation, the binomial probability of a given level being passed with two digital presentations by pure chance equals 0.063. Using a similar paradigm, albeit with two targets per presentation (as is advised in the handbook accompanying KACI), the probability of two consecutive correct identifications at a given level arising by pure chance is 0.25.

With PVi, a change was made to the codebase altering the stair-casing (Fig. 2B): instead of terminating progression down the staircase after one error and retesting the previous level higher in the staircase, further presentations were given at the same level, and two out of three correct test screen presentations (each with four potential positions) were required for a level to be passed.

Our expectation was that the more nuanced staircasing in PVi, allowing for staircase reversals, would increase accuracy. Given the significant differences in population and test builds between studies 1 and 2, direct comparison of repeatability indices must be interpreted with caution, though it is interesting to note that PVb1 exhibited apparent superior indices of repeatability than did PVi. This is likely to relate, in part, to the fewer testable levels at the finer range of acuities with PVb1, contributing to increased clustering around 0.1 and 0.4 levels (Fig. 4B, left panel).

For PVi, the binomial probability of a level being passed with two out of three correct identifications at a given frequency grating, with random target selection, equals 0.141. If two consecutive presentations are correct for a given level (e.g., at the lower end of the staircase), a third presentation at that level is not offered. The probability of these two consecutive presentations being correctly identified by chance equals 0.063 (as with PVb1),

$$P(x) = \frac{N!}{x!(N-x)!} \pi^x (1-\pi)^{N-x},$$

where $P(x)$ = probability of x successes out of N trials, N = number of trials, π = probability of success on a given trial.[12]

The four-target setup in the index tests denotes an intrinsic lower probability of correct levels being passed by chance (0.063–0.141 for PVi and 0.063 PV) than for KACI (0.25), which should theoretically increase reliability.

However, there are several other differences between the paradigms that are likely to impact on reliability. Having four target options on a smaller iPad screen makes refixation eye movement detection more challenging and demands a closer working distance to help mitigate this. In turn, this brings the potential to positively bias toward myopes. Refractive status was not evaluated in either study and would be a desirable aspect of future validation studies.

The closer working distance does bring another advantage, however, in that the screen can be reached by the child at arm's length. In this study, one recognized limitation in the methodology relates to the fact we did not record when children transitioned from a looking response to actively touching the gratings themselves. Although the examiner was not prescriptive regarding looking/pointing/touching behavior, the child was encouraged to tap the touchscreen. The "capture area" around the grating comprised one quarter of the total touchscreen to help allow for minor target-touching "mis-hits" around a given grating. Touching the grating introduces another major divergence from traditional card-based testing, where touching the gratings is discouraged due to the potential for permanently marking/scratching the card. Children were nevertheless encouraged to point to the KACI grating to limit bias. Given the age-groups involved, coordination of hand/arm movement presents another source of potential error, likely to be greater in younger infants and those with concurrent motor impairments.

The typical experience was that in older children their natural tendency was to touch the grating. In less-confident, typically younger children, the test frequently commenced as a preferential looking test, but after a few low-frequency presentations, the behavior often changed to pointing/touching. Recording such behavior in future validation work would allow a more nuanced assessment of the accuracy of the digital platform specifically as a preferential looking test. Furthermore, attention to given behaviors within the context of varying acuity and age subgroups would help evaluate confounding influences.

For use as a preferential looking test, a potential shortcoming of the design of the Peekaboo Vision test screens relates to the difference between vertical and horizontal spacing of the four test grating loci. This presents a potential bias whereby refixations horizontally/diagonally may be relatively easier to spot than vertical refixation eye movements. Indeed, this may exaggerate difficulties in deciding upon looking responses in cases of vertical or horizontal strabismus. Table 2 details the difference in visual angle between the digital platform and Keeler cards. When tested at 25 cm (compared to recommended 38-cm test distance for Keeler Acuity Cards), the center-to-center dis-

tance of the digital grating patches is similar horizontally (1402 vs. 1375 arcmin for Peekaboo Vision), but reduced in the vertical plane to 1045 arcmin, around 25% less than that of the horizontal visual angle between the gratings in Keeler cards. These differences are more pronounced when the digital test distance is extended to the end of the dynamic range to 50 cm, where the vertical visual angle reduces to just 523 arcmin, translating to a refixation eye movement around 37% of the magnitude expected of Keeler cards. While no examiners reported this to be an issue, a potential improvement in the design for future incarnations may be to match horizontal and vertical grating distances from center, or limit the number of loci to two wider spaced target areas, which is an option within the PVi, though not a setting that has been evaluated in the present study. Emerging large display platforms, such as the 12.9 inch iPad Pro[13] provides over 77% increase in display area when compared to the model used in the present study, increasing the scope for incorporation of wider spacing for targets for future versions of digital preferential looking tests.

Where looking responses are replaced with screen tapping or pointing by the child, the issues pertaining to detection of visual behavior are largely obviated. Indeed, for such populations that can reliably tap proximally to the grating patch, there may be advantages to bringing the grating targets closer together such that the gratings fall to a more central retinal position, providing a better index of central macular/foveal acuity.

Another noteworthy difference between the index and reference tests relate to the number of testable acuity levels in each (18 for KACI and PVb1 versus 25 for PVi build), as outlined in Table 1. This difference is also likely to impact on observed reliability differences. Furthermore, the step size in the digital platform becomes coarser toward the finer end of the high-frequency grating range for the digital platforms. By clumping a large *range* of acuities between wide levels in one nominal acuity level, the precision would appear better than when compared with a test that captured more nuanced acuities between steps, but the accuracy may actually be poorer.

With the screen resolution and fixed test distance of 25 cm in study 1, the highest two spatial frequencies possible are created with 1- and 2-pixel grating widths, corresponding with acuity scores of 0.12 and 0.42 logMAR (doubling of visual angle). True acuities lying between 0.42 and 0.12 logMAR were therefore all scored as 0.42 logMAR (Table 1),

resulting in the horizontal clustering of data at 0.42 logMAR in Figure 4B, right panel. In contrast, the Keeler cards with the Children's Additional Set included were −0.1, 0.1, 0.2, and 0.3 logMAR: levels untestable with PVb1. This might be expected to have caused Peekaboo Vision to underestimate acuity in children with good acuities, thereby creating overall disagreement between Peekaboo Vision and Keeler cards, but this was not seen, perhaps because the numbers affected were small. For the PVi version (study 2), the adjustable test distance increased the range of measurable thresholds at the better acuity end of the test, but the differences noted between Peekaboo Vision and Keeler cards remained very small, suggesting there was little or no skew effect with PVb1 in study 1.

In both studies, an animated smiley face comprised the reward animation, and during testing, after the animation reward was demonstrated, the connection between tapping the grating and the subsequent face animation was reinforced to encourage the children to engage. Children with very poor vision, including those with central scotomas, may not have appreciated the details of the smiley face (within the grating in study 1 or within the animation reward in both studies), creating potential confusion. We did not observe any instances where this created an obvious barrier to testing, as even in those children with very poor vision, orientation of attention toward the lowest frequency gratings on an otherwise featureless screen appeared instinctive. It should be noted, however, that statistical analysis regarding acuity was confined to patients who were deemed reliable in reaching an endpoint with all tests (engagement score 2). In our data, this subset captured children with reasonable vision, with only one child exhibiting acuity poorer than 1.0 logMAR. Consequently, impact of a child's inability to detect the face features (in PBb1 or reward in both study 1 and 2) may be a factor missed in the present analysis. Future studies should evaluate Peekaboo Vision in children with severe visual loss, as it is not clear whether the present findings are generalizable.

Bittner et al.[14] compared a digital gratings test (measuring up to 2.2 logMAR), which also employed a four-target forced choice paradigm, comparing instead with the Early Treatment Diabetic Retinopathy Study (ETDRS) as the gold standard in a population of adults who had low vision (legally blind) due to retinal disease. It is interesting that they demonstrated good agreement with digital gratings, which scaled similarly to ETDRS in their retinitis

pigmentosa low-vision group, with ± 2 SD within approximately ± 0.4 logMAR steps.

When comparing the index test (PVb1 and PVi) to the reference standard, we observed around 0.4 logMAR difference in 95% of tests, which is similar to that observed when PVb1 was tested in an adult cohort with artificially degraded vision, using +4, +8, or +16 spherical lenses to evaluate performance in low vision.[5]

The edge artefact at the junction between grating and the high-frequency background checkerboard presents a potential extra visual cue to children. This was a noticeable finding in PVb1 at the finest gratings and also at the junction between grating and isoluminant background at the edges of the face elements (Fig. 1A). A similar finding regarding increased visibility in relation to grating edge effect was reported in relation to the Teller Acuity Test,[15] which prompted the use of the white rings with Keeler Acuity Cards for Children.[11] Following this observation with PVb1 during study 1, the white rings were included in PVi to remove this potential cue, and the face details used in PVb1 were removed.

The data presented here suggest Peekaboo Vision has better repeatability than Keeler preferential looking cards: coefficients of repeatability were 0.27 for Peekaboo Vision versus 0.37 for Keeler cards in study 1 and 0.32 for Peekaboo Vision versus 0.42 for Keeler cards in study 2. This is clinically important, particularly when measuring the change in vision of a child over a period of time.

In study 1 (Malawi), children appeared to engage more with Peekaboo Vision than with Keeler cards, while study 2 (United Kingdom) data suggested no difference. This may reflect changes made in the Peekaboo Vision application, such as the removal of the smiley face detail, or it may reflect differences in the populations tested: tablets are less widely available in Malawi than in the United Kingdom, so their relative novelty may have improved engagement. Other factors such as level of vision and ocular/medical conditions may have influenced engagement, but they are difficult to quantify in such heterogeneous groups. Given that each child had to undertake up to 12 acuity tests (BEO, RE, and LE test and retest for both Peekaboo Vision and Keeler cards) plus a 15-minute interval, better engagement might be expected for a single test in typical practice.

Our data suggest a trend toward decreased engagement on retest across all comparisons, which is to be expected in the given age group. This did not, however, appear to reach statistical significance in our analysis in either study for either platform.

A drop in engagement score was also observed when testing monocularly after BEO testing, with mean engagement score dropping from 1.9 to 1.6 ($N = 49$, Mann-Whitney U test, $W = 2668$, adjusted $P = 0.02$). While statistically significant, the clinical significance of this small drop in engagement score is uncertain. It is possible that an intrinsic feature of PVi increases the potential for disinterest, introducing diminishing engagement with repeat testing. Future design adjustments could accelerate progression on the staircase and increase the variety of sounds and animation rewards to maintain engagement through binocular and monocular tests. Improvements in the study methodology are also desirable in future studies, limiting the testing gamut to replicate a test time more typical of a "real-life" clinical setting.

In study 2, we note wider 95% limits of agreement for Keeler test-retest, approaching an octave step, wider than that found in study 1 with the same test. This may relate to several factors, particularly the very different populations, as well intertester variability. The extensive range of testing (12 tests) in such a young population is likely to be the most significant factor and the most likely reason for the observed trend toward decreased ES across all repeat testing.

There are extensive differences between study 1 and study 2, both in population as well as in design of the index tests. Study 1 is a pilot in nature, testing the methodology in advance of study 2 and also informing the development of the formal PVi build. While it is useful to evaluate in broad terms how a digital infant acuity test performs in Malawi, we cannot draw any meaningful conclusions based on comparisons between the two studies due to intrinsic differences between PVb1 and PVi. The next logical study would seek to repeat the methodology with PVi in a similar cohort in Malawi.

Only total session time was noted for study 1. For study 2, the data inclusion form was amended to capture individual test times. The mean time to test (first test, BEO) was over 1 minute shorter for Peekaboo Vision than for Keeler cards (185 vs. 251 s, paired t test, $N = 33$, $P = 0.0021$), that is, 26% shorter. This may be partly due to Peekaboo Vision's four-choice paradigm compared with Keeler cards' two-choice paradigm. Shorter test time represents an important potential benefit of Peekaboo Vision, given the short attention span of this age group; another is the potential cost saving in high-throughput orthoptic clinics or screening programs.

Compared to card-based vision tests, tablet-based tests are susceptible to veiling glare and to reflections, preventing use outside. On the other hand, a tablet's Lambertian surface can maintain contrast even if the viewing angle is not perpendicular, and photometric compliance of tablets with British and European Standards is at least as good as gold standard retro-illuminated ETDRS charts.[6] Given the potential for variation between reflected light from cards and Lambertian displays, measuring ambient luminance, together with reflected and emitted light from the cards and iPads, respectively, would have been a useful area to explore; however, this was not addressed in the present research, which aimed to evaluate precision and accuracy in a real-life context.

Digital gratings have been used elsewhere in an infant population, using two tablets emulating the Dobson card test.[16] Integration of eye tracking on a monitor-based system has also been evaluated with grating acuity in children[9]; this could increase objectivity and potentially remove the need for a stringent fixed test distance, with live distance measuring between eye and tracker, dynamically adjusting acuity score relative to distance at the moment of refixation.

Furthermore, using a large crowd-sourced data set to train a deep learning convolutional network, the native front-facing mobile camera alone has been demonstrated to predict gaze with an accuracy purported to outperform current state of the art approaches.[17] Combining such technology with the high-fidelity Lambertian tablet display and using gaze to guide preferential looking methodologies through an automated staircase culminates in an exciting possibility whereby visual function could be profiled using a software-only solution on a near ubiquitous mobile platform designed for recreation. Such development could extend the reach of a visual screening program into the patient's home. Indeed, such directions present an exciting new direction for pediatric vision testing, not only for high-contrast acuity, but also for contrast and color assessment.[18,19]

Ongoing regulatory checks of applications for such measures are desirable given the frequent updates to operating systems and hardware. Expansion of national and international standards for vision-testing equipment to include such ubiquitous mobile technology could help support the safe adoption of tablet-based vision testing into regular practice. Further investigation is required to evaluate the role of the technology in amblyopia screening and to evaluate performance in nonexpert testers.

Acknowledgments

Disclosure: **I. Livingstone**, None; **L. Butler**, None; **E. Misanjo**, None; **A. Lok**, None; **D. Middleton**, None; **J.W. Wilson**, None; **S. Delfin**, None; **P. Kayange**, None; **R. Hamilton**; None

References

1. Gilbert C, Foster A. Childhood blindness in the context of VISION 2020—the right to sight. *Bull World Health Organ*. 2001;79:227–232.
2. Maurer D, Lewis TL, Brent HP, et al. Rapid improvement in the acuity of infants after visual input. *Science*. 1999;286:108–110.
3. Daw NW. Critical periods and amblyopia. *Arch Ophthalmol*. 1998;116:502–505.
4. McDonald MA, Dobson V, Sebris SL, et al. The acuity card procedure: a rapid test of infant acuity. *Invest Ophthalmol Vis Sci*. 1985;26:1158–1162.
5. Livingstone IAT, Lok ASL, Tarbert C. New mobile technologies and visual acuity. *Conf Proc IEEE Eng Med Biol Soc*. 2014;2189–2192.
6. Livingstone IAT, Tarbert CM, Giardini ME, et al. Photometric compliance of tablet screens and retro-illuminated acuity charts as visual acuity measurement devices. *PLoS One*. 2016;11: e0150676.

7. Occluding Glasses for Pediatric Vision Testing. Kay Pictures website. Available from: http://kaypictures.co.uk/product/occluding-glasses/. Accessed May 10, 2018.

8. Mayer DL, Beiser AS, Warner AF, et al. Monocular acuity norms for the Teller Acuity Cards between ages one month and four years. *Invest Ophthalmol Vis Sci*. 1995;36:671–685.

9. Jones PR, Kalwarowsky S, Atkinson J, et al. Automated measurement of resolution acuity in infants using remote eye-tracking. *Invest Ophthalmol Vis Sci*. 2014;55:8102–8110.

10. 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.

11. 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.

12. Lane DM. Binomial Distribution. Available from: http://onlinestatbook.com/2/probability/binomial.html. Accessed May 8, 2018.

13. iPad Pro. Technical Specification website. Apple (United Kingdom). Available from: https://www.apple.com/uk/ipad-pro/specs/. Accessed May 9, 2018.

14. Bittner AK, Jeter P, Dagnelie G. Grating acuity and contrast tests for clinical trials of severe vision loss. *Optom Vis Sci*. 2011;88:1153–1163.

15. Robinson J, Moseley MJ, Fielder AR. Grating acuity cards: spurious resolution and the 'edge artifact.' *Clin Vision Sci*. 1998;3:285–288.

16. Mohan KM, Miller JM, Harvey EM, et al. Assessment of grating acuity in infants and toddlers using an electronic acuity card: the Dobson card. *J Pediatr Ophthalmol Strabismus*. 2016;53:56–59.

17. Krafka K, Khosla A, Kellnhofer P, et al. Eye tracking for everyone. Paper presented at the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016, pp.2176–2184. Available from: https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/Krafka_Eye_Tracking_for_CVPR_2016_paper.html

18. Aslam TM, Murray IJ, Lai MYT, et al. An assessment of a modern touch-screen tablet computer with reference to core physical characteristics necessary for clinical vision testing. *J R Soc Interface*. 2013;10:20130239.

19. Tahir HJ, Murray IJ, Parry NRA, et al. Optimisation and assessment of three modern touch screen tablet computers for clinical vision testing. *PLoS One*. 2014;9:e95074.

translational vision science & technology