PEDIATRICS

CrossMark

Validation of dynamic random dot stereotests in pediatric vision screening

Anna Budai[1] · András Czigler[1] · Eszter Mikó-Baráth[1] · Vanda A. Nemes[1] · Gábor Horváth[1] · Ágota Pusztai[2] · David P. Piñero[3] · Gábor Jandó[1]

Abstract

Purpose Stereo vision tests are widely used in the clinical practice for screening amblyopia and amblyogenic conditions. According to literature, none of these tests seems to be suitable to be used alone as a simple and reliable tool. There has been a growing interest in developing new types of stereo vision tests, with sufficient sensitivity to detect amblyopia. This new generation of assessment tools should be computer based, and their reliability must be statistically warranted. The present study reports the clinical evaluation of a screening system based on random dot stereograms using a tablet as display. Specifically, a dynamic random dot stereotest with binocularly detectable Snellen-E optotype (DRDSE) was used and compared with the Lang II stereotest.

Methods A total of 141 children (aged 4–14, mean age 8.9) were examined in a field study at the Department of Ophthalmology, Pécs, Hungary. Inclusion criteria consisted of diagnoses of amblyopia, anisometropia, convergent strabismus, and hyperopia. Children with no ophthalmic pathologies were also enrolled as controls. All subjects went through a regular pediatric ophthalmological examination before proceeding to the DRDSE and Lang II tests.

Results DRDSE and Lang II tests were compared in terms of sensitivity and specificity for different conditions. DRDSE had a 100% sensitivity both for amblyopia ($n = 11$) and convergent strabismus ($n = 21$), as well as a 75% sensitivity for hyperopia ($n = 36$). However, the performance of DRDSE was not statistically significant when screening for anisometropia. On the other hand, Lang II proved to have 81.8% sensitivity for amblyopia, 80.9% for strabismus, and only 52.8% for hyperopia. The specificity of DRDSE was 61.2% for amblyopia, 67.3% for strabismus, and 68.6% for hyperopia, respectively. Conversely, Lang II showed about 10% better specificity, 73.8% for amblyopia, 79.2% for strabismus, and 77.9% for hyperopia.

Conclusions The DRDSE test has a better sensitivity for the detection of conditions such as amblyopia or convergent strabismus compared with Lang II, although with slightly lower specificity. If the specificity could be further improved by optimization of the stimulus parameters, while keeping the sensitivity high, DRDSE would be a promising stereo vision test for screening of amblyopia.

Keywords Amblyopia · Binocular vision · Vision screening · Dynamic random dot stereotest · Lang II

✉ Gábor Jandó
gabor.jando@aok.pte.hu

[1] Institute of Physiology, University of Pécs, Medical School, Szigeti út 12., Pécs 7624, Hungary

[2] Department of Ophthalmology, University of Pécs, Medical School, Rákóczi út 2., Pécs 7623, Hungary

[3] Department of Optics, Pharmacology and Anatomy, University of Alicante, Carr. San Vicente del Raspeig, 03690 San Vicente del Raspeig, Alicante, Spain

Introduction

Amblyopia, commonly known as "lazy eye" is a neurodevelopmental disorder of binocular vision. Due to a mismatch between the perceptual information that reaches the cortex from the two eyes, one of them becomes suppressed. The synaptic connections in the visual cortex are reorganized in such a way that the corrected vision on the side of the suppressed eye suffers a reduction, even though no visible abnormalities are present [1, 2]. The significance of amblyopia in the population is not to be underestimated, the worldwide prevalence varies between 1 and 5% depending on

Springer

many factors [3–8]. Unrecognized amblyopia can lead to various functional deficits of the visual system [9–12]. In addition, its psychosocial impact cannot be neglected either [13–17]. The prevalence of amblyopia is about three times greater in an unscreened population compared with a screened population [18]. In this perspective, the need of a reliable screening method is self-evident.

There are strict recommendations about how states should provide vision screening to all children before the age of 6 [19–21]. This is in accord with the guidelines of the American Academy of Pediatrics that recommends at least yearly visual assessment after the age of 3, and even more often in infancy and early childhood [22–26]. According to the professional guidelines of the Ministry of Health in Hungary, vision screening of children is the competence and duty of pediatricians, school doctors, district nurses, and health visitors, depending on the age of the child [27]. Although vision screening of children is obligatory, unfortunately, it is not performed regularly in the practice [28].

Since amblyopia is originated from abnormal early binocular visual experience due to amblyogenic conditions, it is accompanied with the loss or severe impairment of binocular depth perception [3, 29–31]. Considering the vulnerability of the binocular system, it has long been suggested to use stereotests for screening of amblyopia or conditions potentially leading to amblyopia [32–35]. In most stereotests, dichoptic viewing is required, which can be achieved by using one of the channel separation techniques (e.g., polarized or anaglyph images, column-interleaved displays). Ideally, when using the random dot stereograms, the disparity-coded stereoscopic images are only visible when viewed through the appropriate glasses (i.e., red-green or polarizing glasses). The TNO stereotest uses anaglyphic technique [36, 37], while Randot, the Randot Preschool Stereoacuity, and the Stereofly test (also known as Titmus Fly) use polarizing images [38, 39]. All of them are also suitable for measuring stereoacuity, including the Frisby, which is a real depth stereotest and does not require special glasses [40]. Lang and Lang II stereotests utilize the principle of "panography," essentially a column-interleaved technique, and they do not require glasses either [41, 42].

Most of the conventional stereopsis tests available on the market contain monocular cues for various reasons, which can de-camouflage the cyclopean target [43, 44]. Additionally, all of them have a predetermined set of stimuli displayed on a plastic or paper board. Due to the limited number of figures, the test can be circumvented because motivated children are likely to memorize the expected responses. This effect is even more prominent in a school or kindergarten screening situation when children have an opportunity to communicate with their mates. Both the monocular cues and the predetermined set of figures can affect the ratio of both the false-positive and false-negative passes [45]. Despite the low sensitivity [46], we used the Lang II stereotest as reference in our study, because it

is widespread and available in most clinical settings and family doctor practices in Hungary.

Disadvantages of the recently available stereo vision tests creates the need for developing more versatile tools using new innovative mobile technology that are free from the shortcomings and ease the documentation by using cloud communication (e.g., tablets, smartphones) [44].

Our laboratory has also begun to develop and evaluate a screening system called EuvisionTab, which is a tablet-based Android software, seems to be free of the above mentioned disadvantages, but it has not been tested in the clinical practice yet.

The aim of the present study was to investigate the diagnostic value of EuvisionTab and to compare it with the Lang II in terms of sensitivity and specificity in the detection of amblyopia and other amblyogenic factors [47]. Besides the evaluation of the performance of EuvisionTab, the determination of an optimal threshold for the pass/fail criteria was also the goal of this study. This clinical study was the first report of the results in developing a commercially available and practical test.

Methods

Participants and recruitment

A total of 141 children (aged 4–14, mean age 8.9, SD 2.63) were enrolled in the study for the validation of EuvisionTab. Patients were selected at the pediatric ophthalmology outpatient clinic of the Department of Ophthalmology, University of Pécs, Medical School in Hungary. Inclusion criteria consisted of diagnoses of amblyopia, anisometropia, convergent strabismus, and hyperopia. These were the target conditions. Healthy children or subjects without eye pathologies and with potentially intact stereo vision were also recruited as age-matched controls ($n = 75$). Convenience sampling was used in the selection of these patients. Before the statistical analysis, we had to exclude 19 subjects out of the 141 patients, because of various pathologies, such as Down's syndrome, Marfan syndrome, nystagmus, congenital cataract, retinitis pigmentosa, bilateral congenital cataract, or retinopathy of prematurity. We chose not to include their data in the final analysis because the primary reason of this study was the validation of EuvisionTab as a screening test, and these patients were already under treatment or have been diagnosed beforehand on the grounds of their existing illnesses.

The diagnosis of amblyopia was based on reduced visual acuity (at least 0.8 or worse for the amblyopic eye) despite optimal refractive correction on the otherwise healthy eye [4]. Anisometropia was defined as the difference in refractive error between the two eyes of 1 diopter or more [48]. All other highlighted pathologies such as strabismus [49], and

hyperopia [50], were determined according to the guidelines established in the international literature cited above.

All procedures performed in this study involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This study was approved by the Regional and Local Ethics Committee at the University of Pécs (registration number: 5117). After explaining the parents and children the course of the study and potential risks and benefits, written informed consent was obtained from all parents or legal guardians. Oral informed consent was obtained from all children included in the study.

Stimuli and devices

EuvisionTab (Euvision Ltd., Pécs, Hungary) is a screening system developed at the University of Pécs as a new innovative medical diagnostic tool. Part of the system is a stereo vision module, based on random dot stereogram (RDS) technique for screening of amblyopia. The stereo vision module is essentially an RDS image generator with a great flexibility, because the parameters of the RDSs are freely adjustable. In the present study, we tested a dynamic version of the RDS, which had an embedded, disparity-defined Snellen-E optotype (DRDSE) presented in random orientations (up, down, left, or right). The orientation of the "E" was visible only if the images were viewed binocularly through red-green anaglyph glasses and the observer had intact stereopsis. The red-green glasses contained R26 low-pass (red) and YG09 band-pass (green) gelatin filters (Tobias Optic, Ltd., Budapest, Hungary). More details about the filter characteristics can be found in Markó et al. [51]. The actual screening set comprised a series of nine binocularly visible Snellen E letters and two monocularly visible Snellen-Es. The latter two served as control, to check if the subjects had really understood the task and were able to identify the Snellen E orientation regardless of the existing ocular condition.

We presented our stimuli on a Samsung Galaxy Tab 10.1 (CE0168 with Android 3.2 operating system.). Photometer was used to measure luminance (ILT-1700 Photometer, International Light Technologies, Peabody, USA). The mean luminance and standard deviation of the bright and dark pixels through the color filters were 22.937 cd/m^2 (± 6.21 SD) and 0.72 cd/m^2 (± 0.54 SD), respectively. The disparity and the dot size of the images were 420 sec of arc, the line thickness of the Snellen E optotype subtended 2° of visual angle from the 25- to 30-cm viewing distance. The 420 sec of arc was chosen because the human stereoscopic system is tuned to a certain range of crossed disparities. This disparity is the lowest within the optimum range [52]. The dot density of the images was 1%, with 3% uncorrelated noise. Images were updated at

10 Hz. Before the presentation of the 9 + 2 DRDSE, eight other, easier practice RDS images were shown to become familiar with the task and the stimuli. These RDS images were composed of either static (non-refreshing) dots, or they had larger disparity and a higher dot density. The responses for these practice sessions were not analyzed.

Lang II stereotest (Lang Stereotest AG, Forch, Switzerland) was performed according to the user's manual provided with the test.

Procedures

Clinical examination

Children who visited the pediatric ophthalmology outpatient clinic with their guardian underwent a regular comprehensive eye examination by a trained pediatric ophthalmologist. The examination started with the assessment of visual acuity (VA) of both eyes using an illuminated Snellen E chart from a 5-m distance. They had to use their own pair of spectacles when it was applicable. Cover and cornea light reflex tests were performed using an ophthalmoscope to determine potential misalignment of the eyes. The examination lasted approximately 10 to 12 min/child. The collected data and diagnosis was handed over by the pediatric ophthalmologist and served as reference.

DRDSE testing procedure

After the pediatric ophthalmologist examination, oral consent was acquired from the children and written consent was obtained from the parents or legal guardians. A practice session preceded the actual testing where children could familiarize themselves with the task. Subjects were then seated in a quiet, separate dark room in order to avoid reflections on the screen and to provide better visibility for the stereograms. Children were asked to determine the orientation of each Snellen E optotype, that is, to either tell the direction or indicate it with their hand. We explained the task to the participants and double-checked their understanding with the presentation of non-stereo E letters. The children were given red-green spectacles, which were available in several sizes. The eight practicing figures were presented first and then the actual testing came. Although a short black period indicated the arrival of the new figure, children were also called upon when a new orientation had to be determined. The question, "Now, which way do the legs point?" was always asked for each letter. No suggestions were made for the child. The examiner then entered the response of the child by touching the appropriate button on the screen. Since the protocol was based on a forced choice paradigm, the subjects had to guess even if he/she did not see the letter. Feedback was not provided and the response

time was not limited. The test was automatically terminated after the completion of the 9 + 2 orientations. The first monocularly visible image showed up as the fifth image, while the second as the 11th in the row. The orientations, however, were completely random. The entire procedure lasted for about 5 to 7 min, including parent and patient information, practice with non-stereo E letters, and the examination itself.

In this study, the test was performed by members of the research team, but it could be easily learned and performed by non-professional lay persons as well.

Lang II testing procedure

The Lang II stereotest was presented either before or after the DRDSE in a random order to avoid the failure of the same test due to fatigue or loss of interest.

The plastic card was held upright in front of the subject at 40 cm. Then, they were asked if they could find "the hidden figures" and name them. Since the Lang II has a monocularly visible star, it served as control to test if they understood the task, regardless of binocular pathology, because it is recognizable for patients with no stereopsis as well. If the child was unable to name the objects, we asked them to locate an area on the card where there seemed to be "something interesting" and try to outline its contours, then make a guess of what it could be.

Statistical analysis

Receiving operator characteristic curve

Since we used the same disparity in all test figures, we did not search for a stereo acuity threshold. However, we had to determine how many correct responses were necessary to pass the test. In order to determine the optimal pass/fail criteria for DRDSE and for the evaluation of sensitivity and specificity for detecting the target conditions of the screening test, the "receiver operating characteristic" (ROC) curve analysis was carried out [53]. Thereafter, to determine if detections provided by each stereotest for each ophthalmological condition was consistent or not, we applied the Pearson's chi-square test and Fisher exact test. The threshold for rejecting the null hypothesis was set at $p = 0.001$. Finally, exact values of sensitivity and specificity were calculated according to the standard mathematical formulas [54]. For the statistical analysis, we used SPSS Statistics version 20 (IBM, Armond, NY, USA).

The interpretation of the Lang II test is fairly subjective. According to the official User's Manual for Lang stereotest (Lang Stereotest AG, Forch, Switzerland), the task is to name and show the figures (i.e., monocular: star, binocular: moon (200 sec of arc), car (400 sec of arc), and elephant (600 sec of arc)) on the plastic card. The outcome of the

Lang test is either positive, negative, or doubtful. The test is passed (positive) if all the figures are named and shown correctly. In case of a doubtful answer, the subject should be referred to a specialist for further examination. The result is negative if no object can be detected. These criteria seem to be too strict in the practice, because failure to name the objects correctly can have several reasons: (1) the lack or impairment of stereo vision, (2) the inability to understand the task, and (3) the child is unfamiliar with that object and creatively uses an alternative. Ohlsson et al. previously recognized and described the same methodological problem [55]. They defined normal stereopsis (negative/pass cases) if all the figures were named correctly (although they accepted "fish" for the "car" as normal). Confusing or doubtful answers were considered positive/ fail cases. Some studies carried out in this field accept pointing out two figures out of three [56] or consider the incorrect naming or non-perception of one or more stereoscopic figures as fail [57]. In this specific study, we decided to use the same protocol as recommended by Huynh et al., applied in a study with more than 1700 participants [46]. According to this protocol, the identification of the elephant (disparity, 600 sec of arc) was the pass/fail criterion.

Finally, a comparison between tests (DRDSE vs. Lang) was done for these conditions for which a consistent detection ability was confirmed. For such purpose, the chi-square test was used.

Results

To address the problem of determining the pass/fail criteria for the DRDSE, the threshold was ascertained according to the ROC curve (Fig. 1). ROC curves are shown for the following conditions: hyperopia, convergent strabismus, amblyopia, and anisometropia. Furthermore, we established a fifth group for the subjects who had any of the aforementioned conditions (named "all conditions").

Detailed ROC analysis conducted for the DRDSE test is shown in Table 1, including the area under the curve (AUC),

Fig. 1 Receiver operating characteristic curves for hyperopia (**a**), ▶ convergent strabismus (**b**), amblyopia (**c**), anisometropia (**d**), and all conditions (**e**). The blue line plots the true-positive rate (sensitivity) as a function of the false-positive rate (1-specificity). The distance of the ROC curve (blue line) to the upper left corner indicates the overall accuracy level of the test [73]. Points below the orange line represent worse than random results. Note that the ROC curve of anisometropia (**d**) falls below the orange diagonal (random guess), which divides the ROC space, but in all the other conditions (**a**, **b**, **c**, **e**), the curves are above the orange diagonal. For exact "area-under-curve" values, see Table 1
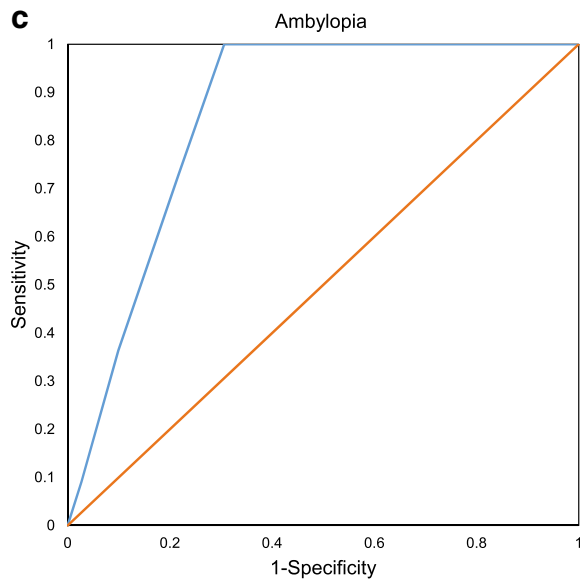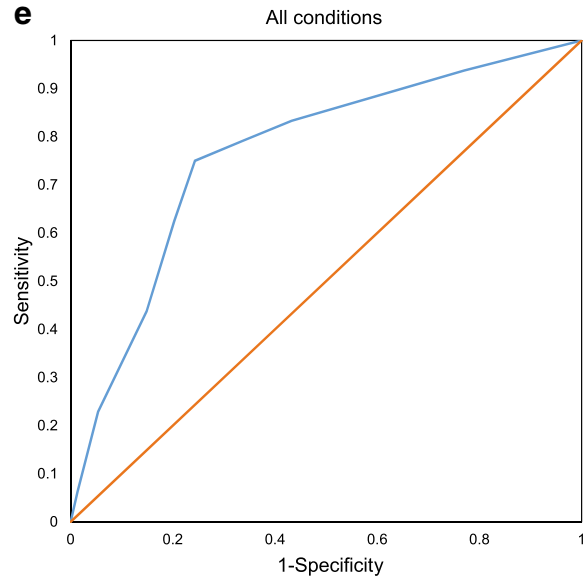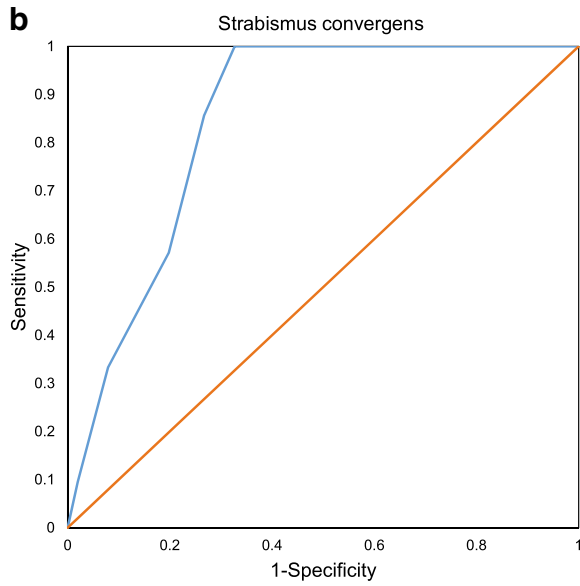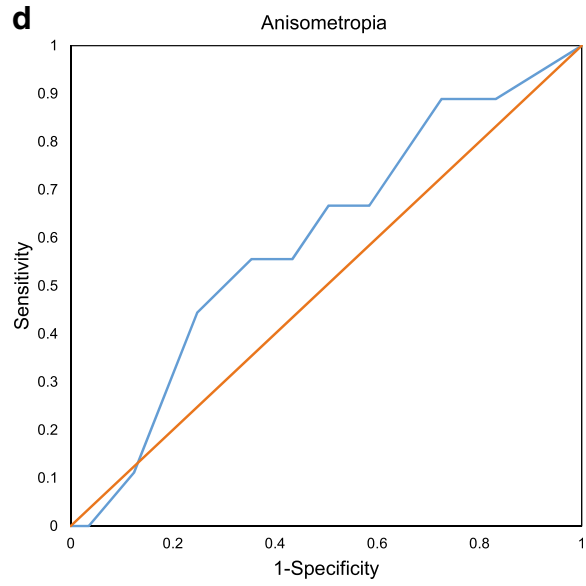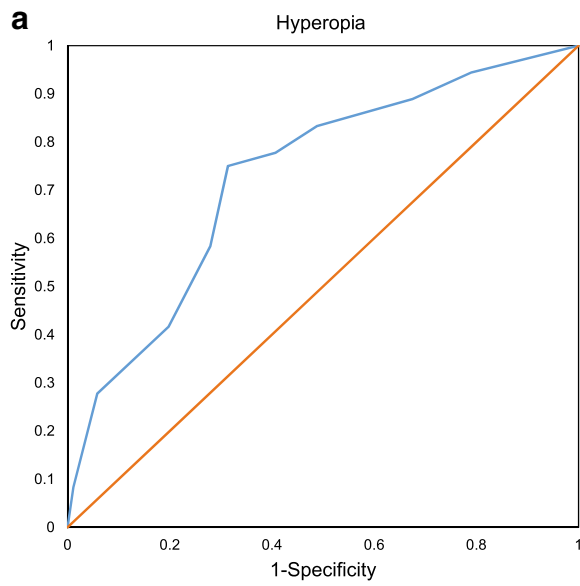
Table 1 Summary table of ROC curve analysis

Diagnoses	Cut-off points (correct responses out of 9)	AUC values (SE)	AUC significance (p value)	Sensitivity (%)	Specificity (%)
Hyperopia	5	0.72 (0.05)	< 0.001*	75	68.6
Hyperopia	6	0.72 (0.05)	< 0.001*	77.8	59.3
Conv.strab.	5	0.84 (0.035)	< 0.001*	100	67.3
Amblyopia	4	0.85 (0.038)	< 0.001*	100	69.4
Amblyopia	5	0.85 (0.038)	< 0.001*	100	61.2
Anisometropia	4	0.59 (0.095)	0.337	55.6	64.6
All conditions	5	0.76 (0.045)	< 0.001*	75	75.6
All conditions	6	0.76 (0.045)	< 0.001*	79.2	66.2

For the categories hyperopia, amblyopia, and all conditions, two thresholds were investigated, in terms of which threshold would provide the highest sensitivity with the best possible specificity

*Significance level

cut-off points, sensitivity and specificity, standard error, and p values (Table 1).

As indicated above, the AUC for the detection ability of DRDSE test was statistically significant for all the target conditions except anisometropia (AUC = 0.59; $p = 0.337$). The area under the ROC curve quantifies the overall ability of the test to discriminate between individuals with or without the disease, and for this matter DRDSE is not a suitable test for the screening anisometropia. When evaluating the cut-off points in different diseases to achieve the best sensitivity and specificity ratio, we found that the selection of different cut-off points was necessary. As an example, for amblyopes, the pass level could be considered at 4/9, which was associated to a sensitivity of 100% and a specificity of 69.4%. In contrast, for hyperopes, a threshold of 5/9 was found to be optimal. Table 2 summarizes the number of cases identified either true positive or true negative when using the 5/9 as a pass/fail threshold of DRDSE test. An additional reason for rejecting pass level of 4/9 is that it would have been accompanied by too high probability ($p = 0.16$) of passing by chance (see Table 4 in the Appendix).

Fisher exact test was performed to evaluate which amblyogenic condition can be significantly detected by Lang II (Table 3).

Lang II stereotest proved to provide significantly consistent detections of amblyopia, strabismus, and "all conditions." However, we could not reject the null hypothesis for anisometropia and hyperopia at our predetermined $p = 0.001$ threshold. When the overall performance was compared with the DRDSE (Fig. 2), anisometropia was excluded from the comparison, since none of the tests detect this condition consistently. Figure 2 represents sensitivity (a) and specificity (b) values of the two stereotests under investigation.

Sensitivity and specificity of DRDSE and Lang II tests

Chi-square test was carried out on the target population ("all conditions," $n = 51$) to determine whether the pass/fail ratio of the two tests were statistically significant and whether they could be used interchangeably. The p value was 0.035, which means we can keep the null hypothesis at our predetermined ($p = 0.001$) significance level.

Table 2 True-positive and true-negative cases for DRDSE using the pass level of 5

Diagnoses	Number of cases (n)	True positive (TP)	True negative (TN)	Significance (p value)	Sensitivity (%)	Specificity (%)
Hyperopia	36	27	59	< 0.0001*	75	68.6
Convergent strabismus	21	21	68	< 0.0001*	100	67.3
Amblyopia	11	11	68	< 0.0001*	100	61.2
Anisometropia	9	5	64	0.337	55.6	56.6
All conditions	48	36	56	< 0.0001*	75	75.6

Number of cases indicate the patients who had the specific diagnosis. True positives (TP) were the affected subjects correctly identified by the test, whereas true negatives (TN) are the ones correctly identified as healthy [74]

*Indicates significance—with this cut-off value, DRDSE proved to be significant as a screening test for hyperopia, convergent strabismus, amblyopia and all conditions, in other words, patients with these diagnoses are likely to fail DRDSE; sensitivity and specificity are shown in percentage

Table 3 True-positive and true-negative cases for the Lang II test and the *p* values for each target condition

| Diagnoses	Number of cases (*n*)	True positive (TP)	True negative (TN)	Significance (*p* value)	Sensitivity (%)	Specificity (%)
Hyperopia	36	19	67	0.0013	52.7	77.9
Convergent strabismus	21	17	80	< 0.00001*	80.9	79.2
Amblyopia	11	9	82	0.0004*	81.8	73.8
Anisometropia	9	6	81	0.004	85.7	71.6
All conditions	48	29	62	< 0.00001*	56.8	87.3

*Indicates significance

Discussion

This is the first study, which reports data about the clinical performance of DRDSE stereovision test, which is a product of a new innovative mobile technology, developed for mass screening of large population of children. This type of screening helps to identify amblyopia and amblyogenic conditions

efficiently. The most important findings are as follows: (1) the DRDSE test significantly detects amblyopia, convergent strabismus and hyperopia, but fails to detect anisometropia; (2) the sensitivity of DRDSE for amblyopia and convergent strabismus was 100%; (3) the overall sensitivity and specificity, which includes target conditions of amblyopia, convergent strabismus, anisometropia, and hyperopia, were 75.0 and

Fig. 2 Bar charts for the comparison of the sensitivity (**a**) and specificity (**b**) values of the Lang II test (blue bars) and the DRDSE (red bars). Find exact percent values on the top of each bar
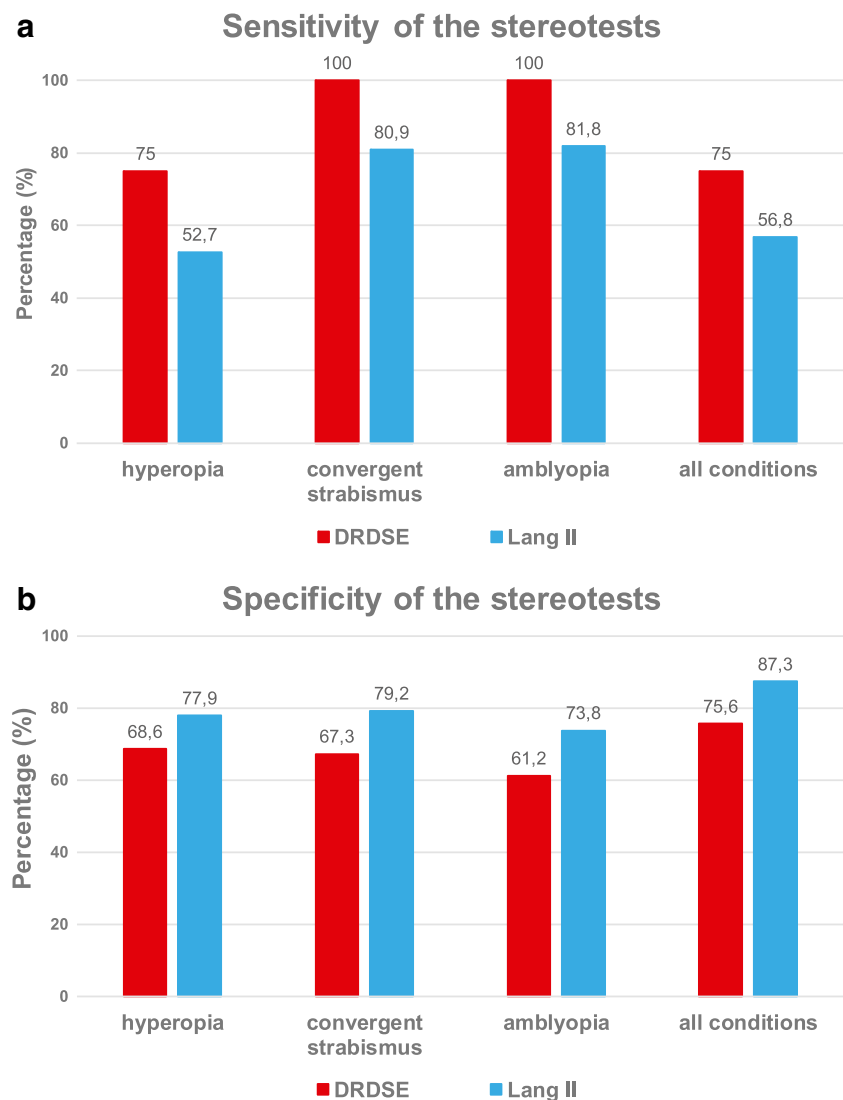
75.6%, respectively; and (4) the overall sensitivity of Lang II test was 56.8%, which is less than that of the DRDSE; however, specificity was 87.3%, which is better than that of the DRDSE.

The present study supports the previously published data on the use of dynamic stereotests in screening [49, 58, 59]. The notion that standard stereotests such as Lang II, Frisby, TNO, etc. are simple and reliable tools for the detection of amblyogenic conditions has been questioned by many authors [57, 60, 61], and it supports the need for better tests. The ideal test is quick, easy to use, and has high sensitivity and specificity at the same time. Sensitivity is the most important parameter among the measures of performance of a screening method. Poor sensitivity, which is due to the high rate of "false passes," increases the number of unrecognized pathologic conditions, which is ethically unaffordable. On the other hand, lower specificity causes overreferral, which is costly for healthcare system. Although both measures are important for the overall efficacy, poorer specificity could be forgiven but no compromise is acceptable in sensitivity. Many attempts have been made to investigate the efficacy of stereotests in vision screening, with the idea that they could simplify and speed up the examination. The additional information about the stereopsis status could modify the referral protocol, which in practice mostly relies on visual acuity. This practice can be originated from the most common definition of amblyopia which emphasizes the decrease in monocular visual acuity [4], despite the fact that the decreased visual acuity is a direct consequence of abnormal development of binocular vision [62]. It has been demonstrated in large-scale screenings that testing stereopsis could be faster and cognitively easier than visual acuity measurements [63]. Although some studies draw attention to the finding that stereotests with a low specificity and sensitivity are not recommended alone for vision screening [64, 65]. Besides, stereotests may fail to identify

conditions where there is no significant loss in binocular vision, such as myopia [66], although symmetric myopia alone is not a typical amblyogenic condition. It is well known that visual acuity test by itself gives overreferrals, requires trained personnel, and can be time consuming and tedious for the children [67]. This controversy could potentially be resolved by constructing a stereotest with the highest possible sensitivity and specificity.

The sensitivity of the DRDSE test was 100%. To date, this is the highest sensitivity stereotest among the available ones.

Next, we will try to find explanations why the sensitivity can be so excellent compared with other tests, why the specificity is worse than for the Lang II, and how could we improve specificity while maintaining the sensitivity.

The excellent sensitivity can be explained by the level of difficulty. Low-density RDS requires global stereopsis to recognize the embedded disparity coded images. The RDS density was 1% with 3% uncorrelated noise, which is probably close to the detection threshold for children. The set of parameters which was used in the protocol has a very low chance for monocular artifact (Budai 2012, unpublished results). In addition to low dot density, the test was dynamic with a refresh rate of 10 Hz. The dynamically refreshing dots may add more difficulty in the perception of the disparity-coded image.

The high sensitivity, which is a considerable advantage of the stimulus might go hand in hand with its disadvantage (i.e., low specificity): it is a rather a hard task even for emmetropes to accomplish. Since the pass threshold of DRDSE has been optimized carefully and it cannot be further lowered, in order to diminish the level of difficulty, we may need to design different screening sets with increased dot density and/or reduced noise level, in favor of elevating the specificity of DRDSE, while preserving high sensitivity.

Younger patients are more likely to fail DRDSE due to their less developed intellect, even if they do not have amblyogenic

Fig. 3 Scatter plot shows the distribution of correct responses in relation to age (years)


Ratio of age and correct responses

factors. In our study we have found a great number of false positive cases ($n = 18$, mean age 8.8 years, SD 2.6). These are the children who did not have any of the highlighted pathologies but failed the test (Fig. 3). A possible explanation for this would have been the lower age of these patients, but Spearman's rank correlation showed no significant positive relation between age and the performance on the test. Therefore, younger age cannot be accounted for false positivity.

Co-existing but potentially non-amblyogenic ophthalmologic conditions may result in decreased stereopsis and increase the number false positive cases. The 18 false positive patients had various eye conditions which are represented in Fig. 4. Astigmatism ($n = 6$) and myopia ($n = 8$) were the most frequently occurring refractive errors. We calculated a mean diopter for myopic patients (o.d. mean, $-2.1D$ and SD, 1.8; o.s. mean, $-2.0D$ and SD, 1.9). Concerning this information, we suspect that a certain degree of myopia alone or combined with astigmatism might lead to a decrease in stereovision. The effect of astigmatism on stereopsis is especially severe when the difference of the angles between the two eyes is more than 45° [68]. In our dataset, two patients had 40° or greater difference. This is in accord with the conclusions of Yang et al. and Kulkarni et al., who demonstrated a correlation between myopia, astigmatism, and reduced stereopsis [69, 70].

The comparison of DRDSE and Lang II is problematic, because of the subjective interpretation [46, 55] (it has already been mentioned in the "Methods" section) and its qualitative result. Due to the lack of quantitative measure, the ROC analysis of Lang II is impossible, therefore optimization of pass/fail threshold was not feasible; these differences in the tests presumably influence the sensitivity and specificity.

Anisometropia was the least detectable condition with DRDSE ($AUC = 0.59$); this observation was also pointed out by Afsari et al. in their study comparing the diagnostic reliability of various stereoacuity tests [71]. It should be
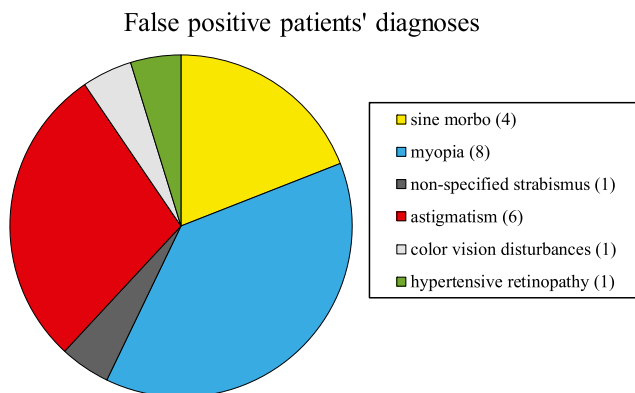
considered that relatively good levels of stereopsis can be present in patients with low levels of anisometropia [72].

The low sensitivity of Lang II is not surprising; this is in accordance with other investigators who found that the sensitivity of Lang stereotest is low and varies between 31.6 and 40% [57, 61]. The comparative analysis of pass/fail ratio in the "all-condition" population suggests that DRDSE and Lang tests are showing similar results, hence they could be used interchangeably.

In this study, we have found that the DRDSE test can be completed in about 5–7 min (including the explanation of the task to the children) and is easy to use (after a short training) for non-professional examiners such as teachers, parents, district nurses, or social workers. As for the technical background, a tablet with the appropriate software and red-green spectacles are required.

Our results are encouraging and should be validated by a larger sample size on the target population (4–6 years of age) for the screening instead of using the convenient sampling method on wider age group. Further works need to be done to improve the relatively low specificity.

Compliance with ethical standards

Conflict of interest Development of the EuvisionTab software is supported by the innovation program of the University of Pécs. Authors Gábor Jandó and Eszter Mikó-Baráth are owners of Euvision Ltd., which is planning to commercialize this product in the future with license from University of Pécs.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

False positive patients' diagnoses

- sine morbo (4)
- myopia (8)
- non-specified strabismus (1)
- astigmatism (6)
- color vision disturbances (1)
- hypertensive retinopathy (1)

Fig. 4 Ophthalmological diagnoses of false positive patients. Numbers in brackets show the number of cases. Some patients had more than one condition

Appendix. Random error of the pass/fail threshold

Any vision test, which is based on discrete number of forced choice selections can be passed by pure mathematical chance even if the subject does not look at the test. We

Table 4 The probability of passing the test by blind guess for different pass levels

Pass level	Probability (p)
1	0.92
2	0.69
3	0.39
4	0.16
5	0.048
6	0.0099
7	0.0013
8	0.0001
9	0.0000038

call this "blind guess." The probability of a certain outcome of a test result can be calculated by using the Bernoulli's formula called "probability ratio." The probability of pass by blind guess can be calculated by summing the probability at the pass level and adding all the probabilities of the better than pass level outcomes:

$$p = \sum_{i=0}^{r-c} \frac{(n-1)^i r!}{i!(r-i)!n^r}$$

where p is the probability of passing the test by blind guess, n is the number of alternative orientations (i.e., $n = 4$), r is the number of trials presented in the test (i.e., $r = 9$), c is the pass level (number of minimum correct responses). Table 4 shows these probabilities at different pass levels.

References

1. Friendly DS (1987) Amblyopia: definition, classification, diagnosis, and management considerations for pediatricians, family physicians, and general practitioners. Pediatr Clin N Am 34(6):1389–1401
2. Ciuffreda KJ, Levi DM, Selenow A (1991) Amblyopia: basic and clinical aspects. Butterworth-Heinemann
3. Webber AL, Wood J (2005) Amblyopia: prevalence, natural history, functional effects and treatment. Clin Exp Optom 88(6):365–375
4. Holmes JM, Clarke MP (2006) Amblyopia. Lancet 367(9519):1343–1351
5. Attebo K, Mitchell P, Cumming R, Smith W, Jolly N, Sparkes R (1998) Prevalence and causes of amblyopia in an adult population. Ophthalmology 105(1):154–159
6. Brown SA, Weih LM, Fu CL, Dimitrov P, Taylor HR, McCarty CA (2000) Prevalence of amblyopia and associated refractive errors in an adult population in Victoria, Australia. Ophthalmic Epidemiol 7(4):249–258
7. McKean-Cowdin R, Cotter SA, Tarczy-Hornoch K, Wen G, Kim J, Borchert M, Varma R, Group M-EPEDS (2013) Prevalence of amblyopia or strabismus in Asian and non-Hispanic white preschool children: multi-ethnic pediatric eye disease study. Ophthalmology 120(10):2117–2124
8. Pan C-W, Chen X, Zhu H, Fu Z, Zhong H, Li J, Huang D, Liu H (2017) School-based assessment of amblyopia and strabismus among multiethnic children in rural China. Sci Rep 7(1):13410
9. Kovács I, Polat U, Pennefather PM, Chandna A, Norcia AM (2000) A new test of contour integration deficits in patients with a history of disrupted binocular experience during visual development. Vis Res 40(13):1775–1783
10. Chandna A, Pennefather PM, Kovács I, Norcia AM (2001) Contour integration deficits in anisometropic amblyopia. Invest Ophthalmol Vis Sci 42(3):875–878
11. Lv B, He H, Li X, Zhang Z, Huang W, Li M, Lu G (2008) Structural and functional deficits in human amblyopia. Neurosci Lett 437(1):5–9
12. Astle AT, McGraw PV (2016) Amblyopia: past, present and future. Ophthalmic Physiol Opt 36(6):611–614
13. Packwood EA, Cruz OA, Rychwalski PJ, Keech RV (1999) The psychosocial effects of amblyopia study. J AAPOS 3(1):15–17
14. Webber AL, Wood JM, Gole GA, Brown B (2008) Effect of amblyopia on self-esteem in children. Optom Vis Sci 85(11):1074–1081
15. Koklanis K, Abel LA, Aroni R (2006) Psychosocial impact of amblyopia and its treatment: a multidisciplinary study. Clin Exp Ophthalmol 34(8):743–750
16. Grant S, Melmoth DR, Morgan MJ, Finlay AL (2007) Prehension deficits in amblyopia. Invest Ophthalmol Vis Sci 48(3):1139–1148
17. Carlton J, Kaltenthaler E (2011) Amblyopia and quality of life: a systematic review. Eye 25(4):403
18. Polling J-R, Loudon SE, Klaver CC (2012) Prevalence of amblyopia and refractive errors in an unscreened population of children. Optom Vis Sci 89(11):e44–e49
19. Bishop A (1991) Vision screening of children: a review of methods and personnel involved within the UK. Ophthalmic Physiol Opt 11(1):3–9
20. Shepard L, Kagan SL, Wurtz E (1998) Principles and recommendations for early childhood assessments
21. Käsmann-Kellner B, Heine M, Pfau B, Singer A, Ruprecht KW (1998) Screening-Untersuchung auf Amblyopie, Strabismus und Refraktionsanomalie bei 1030 Kindergartenkindern. Klin Monatsbl Augenheilkd 213(09):166–173
22. Adolescence IECMC (2008) Recommendations for preventive pediatric health care. Pediatrics 2007:2901
23. Alley CL (2013) Preschool vision screening: update on guidelines and techniques. Curr Opin Ophthalmol 24(5):415–420
24. Hered RW, Wood DL (2013) Preschool vision screening in primary care pediatric practice. Public Health Rep 128(3):189–197
25. Sloot F, Hoeve HL, De Kroon ML, Goedegebure A, Carlton J, Griffiths HJ, Simonsz HJ, Group ERs (2015) Inventory of current EU paediatric vision and hearing screening programmes. J Med Screen 22(2):55–64
26. Cotter SA, Cyert LA, Miller JM, Quinn GE (2015) Vision screening for children 36 to< 72 months: recommended practices. Optom Vis Sci 92(1):6
27. Minisztérium E (2009) Az Egészségügyi Minisztérium szakmai irányelve a 0–18 éves életkorú gyermekek látásfejlődésének követése, a kancsalság és a fénytörési hibák felismeréséről szűrővizsgálati irányelvek védőnők és gyermekorvosok számára Egészségügyi Közlöny, vol 59. Magyar Közlöny Lap-és Könyvkiadó, Budapest
28. Matta NS, Silbert DI (2012) Vision screening across the world. Am Orthopt J 62(1):87–89
29. Donzis PB, Rappazzo JA, Bürde RM, Gordon M (1983) Effect of binocular variations of Snellen's visual acuity on Titmus stereoacuity. Arch Ophthalmol 101(6):930–932

30. Holopigian K, Blake R, Greenwald MJ (1986) Selective losses in binocular vision in anisometropic amblyopes. Vis Res 26(4):621–630

31. McKee SP, Levi DM, Movshon JA (2003) The pattern of visual deficits in amblyopia. J Vis 3(5):5–5

32. Julesz B, Kropfl W, Petrig B (1980) Large evoked potentials to dynamic random-dot correlograms and stereograms permit quick determination of stereopsis. Proc Natl Acad Sci 77(4):2348–2351

33. Simons K, Moss A (1981) A dynamic random dot stereogram-based system for strabismus and amblyopia screening of infants and young children. Comput Biol Med 11(1):33–46

34. Ehrlich MI, Reinecke RD, Simons K (1983) Preschool vision screening for amblyopia and strabismus. Programs, methods, guidelines, 1983. Surv Ophthalmol 28(3):145–163

35. Newman DK, Hitchcock A, McCarthy H, Keast-Butler J, Moore AT (1996) Preschool vision screening: outcome of children referred to the hospital eye service. Br J Ophthalmol 80(12):1077–1082

36. Walraven J (1975) Amblyopia screening with random-dot stereograms. Am J Ophthalmol 80(5):893–900

37. Walraven J, Janzen P (1993) TNO stereopsis test as an aid to the prevention of amblyopia. Ophthalmic Physiol Opt 13(4):350–356

38. Birch E, Williams C, Hunter J, Lapa MC (1997) Random dot stereoacuity of preschool children. J Pediatr Ophthalmol Strabismus 34(4):217–222

39. Fawcett SL, Birch EE (2003) Validity of the Titmus and Randot circles tasks in children with known binocular vision disorders. J AAPOS 7(5):333–338

40. Frisby J (2015) The frisby stereotest: an introduction and a review. http://frisbystereotestcouk/wp-content/uploads/2015/09/Optician-Article-Draft-Low-Resolution-14apr15pdf. Accessed 14 Dec 2017

41. Lang J (1983) A new stereotest. J Pediatr Ophthalmol Strabismus 20(2):72–74

42. Lang J (1984) The two-pencil test and the new Lang stereotest. Br Orthopt J 41:15–21

43. Hahn E, Comstock D, Durling S, MacCarron J, Mulla S, James P, LaRoche R (2010) Monocular clues in seven stereotests. Dalhousie Med J 37(1)

44. Serrano-Pedraza I, Vancleef K, Read JC (2016) Avoiding monocular artifacts in clinical stereotests presented on column-interleaved digital stereoscopic displays. J Vis 16(14):13–13

45. Farvardin M, Afarid M (2007) Evaluation of stereo tests for screening of amblyopia. Iran Red Crescent Med J 9(2):5

46. Huynh SC, Ojaimi E, Robaei D, Rose K, Mitchell P (2005) Accuracy of the Lang II stereotest in screening for binocular disorders in 6-year-old children. Am J Ophthalmol 140(6):1130–1132

47. Birch EE (2013) Amblyopia and binocular vision. Prog Retin Eye Res 33:67–84

48. Ohlsson J, Baumann M, Sjöstrand J, Abrahamsson M (2002) Long term visual outcome in amblyopia treatment. Br J Ophthalmol 86(10):1148–1151

49. Von NG (1990) Binocular vision and ocular motility: theory and management of strabismus

50. Castagno VD, Fassa AG, Carret MLV, Vilela MAP, Meucci RD (2014) Hyperopia: a meta-analysis of prevalence and a review of associated factors among school-aged children. BMC Ophthalmol 14(1):163

51. Markó K, Kiss HJ, Mikó-Baráth E, Bártfai O, Török B, Kovács I, Jandó G (2009) Contrast independence of dynamic random dot correlogram evoked VEP amplitude. J Vis 9(4):8–8

52. Horvath G, Nemes VA, Rado J, Czigler A, Torok B, Buzas P, Jando G (2018) Simple reaction times to cyclopean stimuli reveal that the binocular system is tuned to react faster to near than to far objects. PLoS One 13(1):e0188895. https://doi.org/10.1371/journal.pone.0188895

53. Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P (2004) Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. Am J Epidemiol 159(9):882–890

54. Parikh R, Mathai A, Parikh S, Sekhar GC, Thomas R (2008) Understanding and using sensitivity, specificity and predictive values. Indian J Ophthalmol 56(1):45

55. Ohlsson J, Villarreal G, Sjöström A, Abrahamsson M, Sjöstrand J (2002) Screening for amblyopia and strabismus with the Lang II stereo card. Acta Ophthalmol 80(2):163–166

56. Rasmussen F, Thoren K, Caines E, Andersson J, Tynelius P (2000) Suitability of the Lang II random dot stereotest for detecting manifest strabismus in 3-year-old children at child health centres in Sweden. Acta Paediatr 89(7):824–829

57. Ohlsson J, Villarreal G, Abrahamsson M, Cavazos H, Sjöström A, Sjöstrand J (2001) Screening merits of the Lang II, Frisby, Randot, Titmus, and TNO stereo tests. J AAPOS 5(5):316–322

58. Broadbent H, Westall C (1990) An evaluation of techniques for measuring stereopsis in infants and young children. Ophthalmic Physiol Opt 10(1):3–7

59. Leat SJ, St Pierre J, Hassan-Abadi S, Faubert J (2001) The moving dynamic random dot stereosize test: development, age norms, and comparison with the Frisby, Randot, and Stereo Smile tests. J Pediatr Ophthalmol Strabismus 38(5):284–294

60. Marsh WR, Rawlings SC, Mumma JV (1980) Evaluation of clinical stereoacuity tests. Ophthalmology 87(12):1265–1272

61. Schmidt PP, Kulp M (1994) Detecting ocular and visual anomalies in a vision screening setting using the Lang stereotest. J Am Optom Assoc 65(10):725–731

62. Wiesel TN, Hubel DH (1963) Effects of visual deprivation on morphology and physiology of cells in the cat's lateral geniculate body. J Neurophysiol 26(6):978–993

63. Ruttum MS, Bence SM, Alcorn D (1986) Stereopsis testing in a preschool vision screening program. J Pediatr Ophthalmol Strabismus 23(6):298–302

64. Ruttum MND (1991) Stereopsis testing to reduce overreferral in preschool vision screening. J Pediatr Ophthalmol Strabismus 28:131–133. https://doi.org/10.3928/0191-3913-19910501-04

65. Simons K, Reinecke RD (1974) A reconsideration of amblyopia screening and stereopsis. Am J Ophthalmol 78(4):707–713

66. Hope C, Maslin K (1990) Random dot stereogram E in vision screening of children. Clin Exp Ophthalmol 18(3):319–324

67. Mølgaard IL, Biering-Sørensen K, Michelsen N, Elmer J, Rydberg A (1984) Amblyopia screening in kindergartens with TNO stereotest. Acta Ophthalmol 62(1):156–162

68. Julesz B (1995) Dialogues on perception. MIT Press, Cambridge

69. Yang J-W, Huang T-Y, Lee J-S, Yeung L, Lin Y-F, Sun C-C (2013) Correlation between myopic ametropia and stereoacuity in school-aged children in Taiwan. Jpn J Ophthalmol 57(3):316–319

70. Kulkarni V, Puthran N, Gagal B (2016) Correlation between stereoacuity and experimentally induced graded monocular and binocular astigmatism. J Clin Diagn Res 10(5):NC14

71. Afsari S, Rose KA, Pai AS-I, Gole GA, Leone JF, Burlutsky G, Mitchell P (2013) Diagnostic reliability and normative values of stereoacuity tests in preschool-aged children. Br J Ophthalmol 97(3):308–313

72. G-s Y, Huang J, Maguire MG, Quinn G, Kulp MT, Ciner E, Cyert L, Orel-Bixler D (2013) Associations of anisometropia with unilateral amblyopia, interocular acuity difference, and stereoacuity in preschoolers. Ophthalmology 120(3):495–503

73. Zweig MH, Campbell G (1993) Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin Chem 39(4):561–577

74. Altman DG, Bland JM (1994) Diagnostic tests. 1: sensitivity and specificity. BMJ. Br Med J 308(6943):1552