# *ColourSpot*, a novel gamified tablet-based test for accurate diagnosis of color vision deficiency in young children

Teresa Tang[1] · Leticia Álvaro[2] · James Alvarez[3] · John Maule[1] · Alice Skelton[1] · Anna Franklin[1] · Jenny Bosten[4]

## Abstract

There is a need for a straightforward, accessible and accurate pediatric test for color vision deficiency (CVD). We present and evaluate *ColourSpot*, a self-administered, gamified and color calibrated tablet-based app, which diagnoses CVD from age 4. Children tap colored targets with saturations that are altered adaptively along the three dichromatic confusion lines. Two cohorts (Total, $N = 772$; Discovery, $N = 236$; Validation, $N = 536$) of 4–7-year-old boys were screened using the Ishihara test for Unlettered Persons and the Neitz Test of Color Vision. *ColourSpot* was evaluated by testing any child who made an error on the Ishihara Unlettered test alongside a randomly selected control group who made no errors. Psychometric functions were fit to the data and "threshold ratios" were calculated as the ratio of tritan to protan or deutan thresholds. Based on the threshold ratios derived using an optimal fitting procedure that best categorized children in the discovery cohort, *ColourSpot* showed a sensitivity of 1.00 and a specificity of 0.97 for classifying CVD against the Ishihara Unlettered in the independent validation cohort. *ColourSpot* was also able to categorize individuals with ambiguous results on the Ishihara Unlettered. Compared to the Ishihara Unlettered, the Neitz Test generated an unacceptably high level of false positives. *ColourSpot* is an accurate test for CVD, which could be used by anyone to diagnose CVD in children from the start of their education. *ColourSpot* could also have a wider impact: its interface could be adapted for measuring other aspects of children's visual performance.

**Keywords** Color vision deficiency · Visual development · Mobile health applications · Pediatric · Gamification

## Introduction

Measuring individual visual performance in young children has long been a challenge (Robbins et al., 2003). This has meant that visual problems like color vision deficiency

Teresa Tang and Leticia Álvaro are joint first authors

Anna Franklin and Jenny Bosten are joint senior authors

✉ Anna Franklin
Anna.Franklin@sussex.ac.uk

✉ Jenny Bosten
J.Bosten@sussex.ac.uk

[1] The Sussex Colour Group, School of Psychology, University of Sussex, Falmer, Brighton BN1 9QH, UK

[2] Facultad de Psicología, Universidad Complutense de Madrid, Pozuelo de Alarcón, Madrid 28883, Spain

[3] School of Psychology, University of Sussex, Falmer, Brighton BN1 9QH, UK

[4] Sussex Vision Lab, School of Psychology, University of Sussex, Falmer, Brighton BN1 9QH, UK

(CVD) or other visual abnormalities are often not detected as early as they could be. There is a need to develop widely accessible intuitive psychophysical methods targeted to young children.

Congenital CVD is the most common congenital visual disorder, affecting approximately 8% of males and 0.4% of females (Birch, 2012). In many countries there is no routine screening for CVD in schools or by optometrists (Atowa et al., 2019; Azizoğlu et al., 2017; Ciner et al., 1999; Department of Health, 2009; Holroyd & Hall, 1997; Hopkins et al., 2013; Jadhav et al., 2017; Ramachandran et al., 2014; Stewart-Brown & Haslum, 1988; WHO Programme for the Prevention of Blindness and Deafness, 2003). One of the barriers to mass screening for CVD in children is a lack of widely accessible, high quality and child-appropriate diagnostic tests. However, identifying CVD in young children is necessary to support them in accessing an educational system that relies heavily on color.

Normal color vision involves the comparison of signals sent by three photoreceptor types: short (S), medium (M) and long (L) wavelength-sensitive cones. Anomalous trichromacy is the more common and milder form of red-

green CVD, where all three cone types are present and functional, but there is an abnormality either in the L or in the M cone photopigment's spectral sensitivity. Dichromacy is the more severe form of red-green CVD where there is an absence either of the L or of the M photopigment (Neitz & Neitz, 2000; Parry, 2015; Sharpe et al., 1999). It is important to detect CVD and identify how individuals can be supported, as it can significantly affect quality of life such as health, well-being and work (Barry et al., 2017; Chan et al., 2014; Cole, 2004, 2007; Cumberland et al., 2005; Steward & Cole, 1989; Tagarelli et al., 2004). Ninety percent of adults with CVD encounter problems in their daily lives (Steward & Cole, 1989), while children with CVD are at risk of social, behavioral and emotional difficulties, and adverse educational outcomes (Grassivaro Gallo et al., 1998, 2002; Suero et al., 2005; Thomas et al., 2018; Thuline, 1964). Educational materials, especially those for young children, such as reading schemes or mathematics activities (Birch, 2001), typically rely on color as a learning tool (Rinaldi et al., 2020; Suero et al., 2005), which makes them less accessible for children with CVD. In fact, one simulation of CVD suggests that 10% of tasks in educational textbooks are inaccessible for a child with CVD (Torrents et al., 2011). The dependence on color in education likely means that students with CVD are disadvantaged compared to their peers with normal color vision (Mehta et al., 2018).

Although there are a number of tests which have been used to detect CVD in children, they each have limitations. The gold standard diagnostic test in adults, the anomaloscope, has been used successfully with children older than 7, but the task of mixing a red and green to match a yellow is too demanding for younger children (Verriest, 1982). There are tests which aim to be child-friendly versions of other adult tests for CVD (e.g., the Ishihara test for Unlettered Persons), but their success varies greatly. Table 1 summarizes the properties of existing tests for CVD that have been designed for children or that are presented on tablet displays. An additional table summarizing tests for CVD aimed at adults which have also been used in children is included as Table S1 in the Supplementary Information (SI).

Although Table 1 presents many options for diagnosing CVD in children, no existing test is suitable for mass-screening for CVD from as young as 4 years. Firstly, many of the tests require specialist equipment or resources and/or require a trained administrator. This requires screening of children for CVD to take place in an optometrist's clinic or as part of a well-funded school CVD screening program. The World Health Organization does not currently recommend that color vision screening is included as part of population-based vision assessments (WHO Programme for the Prevention of Blindness and Deafness, 2003), and many countries do not always perform pediatric color vision screenings (Atowa et al., 2019), including Australia (Hopkins et al., 2013),

India (Jadhav et al., 2017), Malaysia (Thomas et al., 2018), Turkey (Azizoğlu et al., 2017), the United Kingdom (Department of Health, 2009) and 80% of states in the United States (Ciner et al., 1999). Secondly, although some tests have good sensitivity and specificity in adults, there is little evidence that they can diagnose CVD accurately in young children. Thirdly, many of the tests are not tailored to the capabilities of young children. For example, pseudoisochromatic plate tests may be difficult if children are not yet sufficiently familiar with the shapes, numbers and animals depicted on the plates (Tekavčič Pompe & Stirn Kranjc, 2012). Children may fail to identify these stimuli for reasons that are unrelated to their color vision. Many of the tests rely on the ability to integrate elements into a holistic percept, a skill which is known to be underdeveloped in young children (Kovács, 2000; Scherf et al., 2009). Some tests have complicated instructions, take too long to complete or are not sufficiently engaging for children's limited attention and motivation.

One potential solution to the limited accessibility of pediatric tests for CVD is to use tablet-based methods. Such methods can provide mass testing at home, school and in the community, and are colorimetrically adequate for color vision testing if properly calibrated (Bodduluri, Boon, & Dain, 2017a; Dain & Almerdef, 2016; de Fez et al., 2016; de Fez, Luque, García-Domene, et al., 2018a). Two iPad-based tests for CVD in children, the DoDo game (Nguyen, Do, et al., 2014a; Nguyen, Lu, et al., 2014b) and the Optopad (de Fez, Luque, Matea, et al., 2018b), have been recently developed. Both tests only partially address the limitations of pediatric tests for CVD listed in Table 1. The DoDo game has the advantage of gamification (Abramov et al., 1984; Bodduluri, Boon, Ryan, & Dain, 2017b; Ling & Dain, 2018), but its calibration process has not been specified, which means that its effectiveness may vary between devices. Optopad is not gamified or specifically designed for children, and therefore may not sufficiently engage children's motivation and attention, both factors that are known to influence children's task performance (Diez et al., 2001; Taylor, 1970). Optopad also requires specialized software and hardware for calibration to be usable on different devices. Furthermore, Optopad identified only 6 out of 341 children as having CVD, a small sample to estimate the test's sensitivity, and a surprisingly small prevalence rate (1.76% compared to an expected prevalence for their sample of 4.69%; Birch, 2012). This low measured prevalence along with low adult sensitivity and specificity (0.75/0.94) and the need to exclude some children (younger participants could enrol only after confirming that they were able to read numbers and correctly identify directions) means that Optopad's child sensitivity and specificity values should be interpreted with caution. Both the DoDo game and Optopad are also not yet readily available for download,

**Table 1**    A summary of tests for CVD either intended for use on children or presented on tablet displays

| Test name | Adult sensitivity/specificity, numbers of adult test participants, and comparison test | Details of tests on children including numbers of participants and sensitivity/specificity where available | Recommended minimum age for test | Limitations |
|---|---|---|---|---|
| **Pseudoisochromatic** | | | | |
| Ishihara for Unlettered Persons* (Ishihara & Ishihara, 1943) | 0.98/1.00 (Birch & McKeever, 1993) (CVD=29, CVN=263) Ishihara, 1989 edition (Ishihara, 1989) | 90% of 3–6-year-old children (N=40) successfully completed the test (Mäntyjärvi et al., 2000) | 4 years (Birch & Platts, 1993) | †, ‡, § Requires shape knowledge and pathway tracing |
| Kojima-Matsubara Color Vision Test plates* (Matsubara & Kojima, 1957) | 0.08/0.90 (Lee et al., 1997) (CVD=13, CVN=20) Anomaloscope | 3–6-year-old children (N=40) successfully completed the test (Mäntyjärvi et al., 2000) | 4 years (Mäntyjärvi et al., 2000) | †, ‡, § Requires animal knowledge |
| Pease Allen Color Test* (Pease & Allen, 1988) | 0.87/1.00 (Pease & Allen, 1988) (CVD=23, CVN=210) Anomaloscope | 97% of 3–6-year-old children passed the test (Pease & Allen, 1988) | 3 years (Pease & Allen, 1988) | †, ‡ |
| Color Vision Testing Made Easy (CVTME)* (Waggoner, 1994) | 0.97/0.90 (Dain, 2010) (CVD=41, CVN=42) Anomaloscope | Children over 4 years successfully completed the test (Richardson et al., 2008) | 3 years (Waggoner, 1994) | †, ‡, § Requires shape, animal and object knowledge |
| Neitz Test of Color Vision* (Neitz & Neitz, 2001) | 1.00/0.86 (Block et al., 2004) (CVD=14, CVN=26) Anomaloscope | Tested in 4–12 years (N=115) and verified with genetic testing (Neitz & Neitz, 2001) | 4 years (Neitz & Neitz, 2001) | †, ‡, § Requires shape knowledge |
| Color Vision Evaluation Test (CVET)* (Fish et al., 2020) | N/A | 3–18 years 0.96/0.96 (Fish et al., 2020) (CVD=70, CVN=85) Ishihara 38-plate edition | 3 years (Fish et al., 2020) | †, § Ability to identify orientations |
| **Oddity** | | | | |
| Mollon-Reffin Minimalist Test* (Mollon et al., 1991) | N/A | 3–10 years successfully completed the test (Shute & Westall, 2000). Children rated as most enjoyable test compared to CVTME, Neitz and Analphabetic Ishihara (Tekavčič Pompe & Stirn Kranjc, 2012) | 3 years (Shute & Westall, 2000) | †, ‡ |
| University of Waterloo Colored Dot Test* (Hovis et al., 2002) | 0.57/1.00 (Hovis et al., 2002) (CVD=21, CVN=31) Anomaloscope | 2.5–5 years 0.48/0.97 (Hovis et al., 2002) (CVD=25, CVN=524) Standard Pseudoisochromatic Plates (Ichikawa et al., 1979; Tanabe et al., 1978) | 3 years (Hovis et al., 2002) | † |
| **Tablet-based** | | | | |
| DoDo game* (Nguyen, Do, et al., 2014; Nguyen, Lu, et al., 2014) | N/A | 6–17 years 0.81/1.00 (Nguyen, Do, et al., 2014) (CVD=16, CVN=16) Ishihara, 1998 edition (Ishihara, 1998) | 2.5 years (Nguyen, Lu, et al., 2014) | † |
| Optopad (de Fez, Luque, Matea, et al., 2018) | 0.75/0.94 (de Fez, Luque, Matea, et al., 2018) (CVD=16, CVN=50) Farnsworth Munsell 100-Hue (Farnsworth, 1943) | 3–11 years 1.00/1.00 (de Fez, Luque, Matea, et al., 2018) (CVD=6, CVN=335) Ishihara, 2002 edition (Ishihara, 2002) | 3 years (de Fez, Luque, Matea, et al., 2018) | †, § Ability to identify orientations of the Landolt C |

The table outlines the test type, adult sensitivity and specificity values with their sample size and comparison test, rates of successful test completion in children with sensitivity and specificity if available, the recommended minimum age for test completion and the limitations of each test

*Note.* Definitions. **Anomaloscope**: The anomaloscope is an optical instrument where individuals are asked to match different mixtures of red and green monochromatic lights to a yellow monochromatic light. It is the gold standard for assessing color vision; **CVD:** Participants with color vision deficiency (any CVD type, e.g., anomalous trichromacy, dichromacy); **CVN:** Participants with normal color vision; **N/A:** Not available; **Pseudoisochromatic tests**: These tests have an array of colored dots that form a figure (digits, pathways, letters, animals, or shapes) against an isoluminant background which individuals are asked to identify; **Oddity tests**: An odd-one-out task where individuals are asked to identify a colored target amongst distractors; **Sensitivity**: The rate at which a diagnostic test identifies true positives (i.e. individuals with a condition are correctly identified). For example, against the comparison test (in this example, the standard Ishihara test), the Ishihara Unlettered has a sensitivity of 0.98, indicating that 98% of individuals are correctly diagnosed as having a CVD (of any type) and 2% are false negatives (i.e., the Ishihara Unlettered diagnosed the individual as having normal color vision (CVN) but the standard Ishihara test diagnosed the same individual as CVD); **Specificity**: The rate at which a diagnostic test identifies true negatives (i.e. correctly identifies the absence of a condition). For example, when compared with the comparison test (in this example, the standard Ishihara test), the Ishihara Unlettered has a specificity of 1.00, indicating that 100% of CVN individuals were correctly categorized as CVN, and 0% of individuals were false positives (i.e., where the Ishihara Unlettered diagnosed the individual as CVD but the standard Ishihara test diagnosed the same individual as CVN).

Symbols. * The test was specifically designed for children. † **Inaccessibility.** The test is inaccessible for the public and/or requires specialized equipment and/or resources and/or a trained specialist administrator. ‡ **Unknown validity**. The sensitivity and specificity values of the tests are unknown in children; § **Unsuitability.** The test requires an understanding of numbers, orientation, shapes, and/or animals, or the task is so demanding that it is unsuitable for young children and children with additional educational needs.

and there is nothing yet published to suggest that they are being used to diagnose CVD in children.

Here we present *ColourSpot,* a novel gamified tablet-based test for CVD in children from 4 years of age, which overcomes the limitations of existing pediatric tests for CVD. Our aim is for *ColourSpot* to be an accessible mass screening tool for CVD, usable by parents, teachers or optometrists in the home, classroom or clinic at the age that children start school. We combine rigorous psychophysics and color calibration with a child-friendly gamified and animated interface. Using an adaptive staircase procedure, *ColourSpot* measures discrimination thresholds for colored targets defined along the three (protan, deutan and tritan) dichromatic color confusion lines (the terms "protan," "deutan" and "tritan" refer to types of CVD relating to the L-, M- and S-cone, respectively; protan, deutan and tritan confusion lines are lines in color space along which colors are not discriminable by dichromats missing the respective cone type). Thresholds are measured in an engaging and simple game where children tap the colored targets among luminance-defined distractors to reveal animated characters.

We outline the procedure and associated accuracy for calibrating display devices, and the psychophysics and design of *ColourSpot.* We present results evaluating *ColourSpot* in a two-stage design that ensures that its validation is based on data from a (validation) cohort independent of the (discovery) cohort used to identify optimal classification parameters for CVD versus normal color vision. Two hundred and thirty-six boys in the discovery cohort and 536 boys in the validation cohort completed the Ishihara test for Unlettered Persons (Ishihara Unlettered) and the Neitz Test of Color Vision (Neitz Test). We were interested in evaluating the latter since it has been proposed as a mass-screening tool for use in classrooms. Any child who made more than one error on the Ishihara Unlettered was tested on *ColourSpot.* A randomly selected control group of children who made no errors on the Ishihara Unlettered was also tested on *ColourSpot.* We present *ColourSpot's* sensitivity and specificity against the Ishihara Unlettered as a pseudo gold standard.

## Methods

### Development of *ColourSpot*

#### Color calibration of iPad screens

In order to render colored stimuli accurately on different models of iPad which vary in their display of color, seven models of iPad (iPad 2, 2011; iPad (3rd generation), 2012; iPad (4th generation), 2012; iPad Air, 2013; iPad Air 2, 2014; iPad Pro 9.7," 2016; and iPad (5th generation), 2017; Apple, Cupertino, CA, USA) were color calibrated. The gamma functions and the spectral power distributions of the

display primaries were measured for a minimum of two units of each model using a PR655 Spectroradiometer (PhotoResearch, Chatsworth, CA, USA). Mean LMS to RGB transformation matrices were created using the radiance spectra of the three primaries, and mean gamma exponents were measured (and corrected). To check the quality of the calibration for each model, we presented our calibrated protan, deutan and tritan stimuli on each iPad model, and measured the spectra of the stimuli, calculating the LMS values of the presented stimuli from the spectra. Calibration was successful, showing relatively small residual systematic errors between the intended and measured chromaticities. The calibration errors tended to be systematic and therefore to place the stimuli on alternative but equally valid confusion lines (see Fig. 1). *ColourSpot* automatically detects the iPad model in use and provides the relevant iPad calibration for that model in order to achieve colorimetric accuracy for the test.

#### Luminance and tritan noise

There are large individual variations in spectral luminosity, particularly for individuals with CVD whose cone spectral sensitivities differ markedly from those of normal trichromats (Judd, 1945; Regan et al., 1994; Teller et al., 2003). To address this, we modeled the maximum luminance signal from our chromatic targets for dichromats and found this to be 10% of the background luminance. We therefore masked this potential signal by drawing the luminance of the targets and distractors from a linear distribution between $\pm$ 20% of the average luminance. Similarly, due to individual differences in cone fundamentals (Bosten, 2019) and small residual color calibration errors, binary tritan noise was added over the stimulus display at $\pm$ 16% of the background S/(L+M) value. This value was chosen after estimating the maximum available tritan signal from the protan and deutan targets (due either to individual differences in the peak sensitivities of the cone fundamentals or to calibration error) at 8%.

#### ColourSpot design

*ColourSpot* measures protan, deutan and tritan discrimination thresholds. On each trial, participants are shown three targets selected along each of the dichromatic confusion axes (one protan, one deutan and one tritan target, see Fig. 1), and eight achromatic distractors varying in luminance (see Fig. 2). The test begins with a tutorial animation where an animated monkey character jumps up and touches a colored target whilst a child's voiceover says, "tap a colored spot." When a colored target is tapped it disappears to reveal a smaller animated character that moves and makes an appropriate animal sound (see Fig. 3). The tutorial uses highly saturated colored targets that are visible for both individuals with normal color vision and individuals with CVD. After the monkey demonstrates
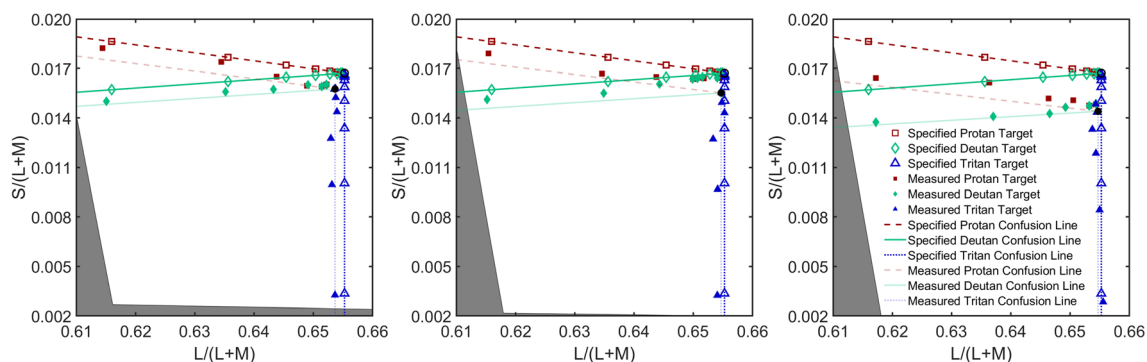
**Fig. 1** Calibration results for iPad Pro 9.7" (left), iPad (5th Generation) (middle) and iPad Air 2 (right) plotted in the MacLeod-Boynton chromaticity diagram (MacLeod & Boynton, 1979). The open symbols show the specified chromaticities and the closed symbols show the measured chromaticities. The red squares are targets on the protan confusion line (dashed line), the green diamonds are targets on the deutan confusion line (solid line) and the blue triangles are the targets on the tritan confusion line (dotted line). Each confusion line passes through the specified (open black circle) or measured (closed black circle) standard D65 white point. Opaque lines represent specified confusion lines (passing through the specified white point) and semi-transparent lines represent the confusion lines closest to the measured stimuli (passing through the measured white point). The solid grey area is outside the iPad's gamut

the rule of the game three times, the child begins practice trials. If the child taps a colored spot successfully five consecutive times in the practice trials, they progress automatically to the main part of the test which measures thresholds for the protan, deutan and tritan targets. On each trial, if a target is tapped, a cartoon animal animation along with an associated sound (e.g., a bird and a chirp) is revealed as a reward. There is no sound or animation if a distractor is tapped. Periodically, child voiceover clips are presented saying "Well done!," "Amazing" and "You're good at this!" as positive reinforcement for engaging with the game. During the main part of the test, there are five animated scene changes to keep the children engaged (Figs. 2 and 3 and supplementary video S1).

To measure thresholds, the colored targets in the main part of the test vary along the protan, deutan and tritan confusion lines according to an adaptive staircase procedure. Each target type (protan, deutan, tritan) has its individual staircase which begins at the highest available saturation level. If a target of a particular type is tapped, its saturation is multiplied by a factor of 0.5 for the next trial, thus decreasing the saturation of targets of that type, making detection of the target type more difficult. If a distractor is tapped, the saturations of all three target types are multiplied by a factor of 1.5 for the next trial, thus making the next trial easier. This design is intended to be helpful for children with severe CVD who may be performing at floor level for protan and deutan targets. Having a tritan target available should reduce their
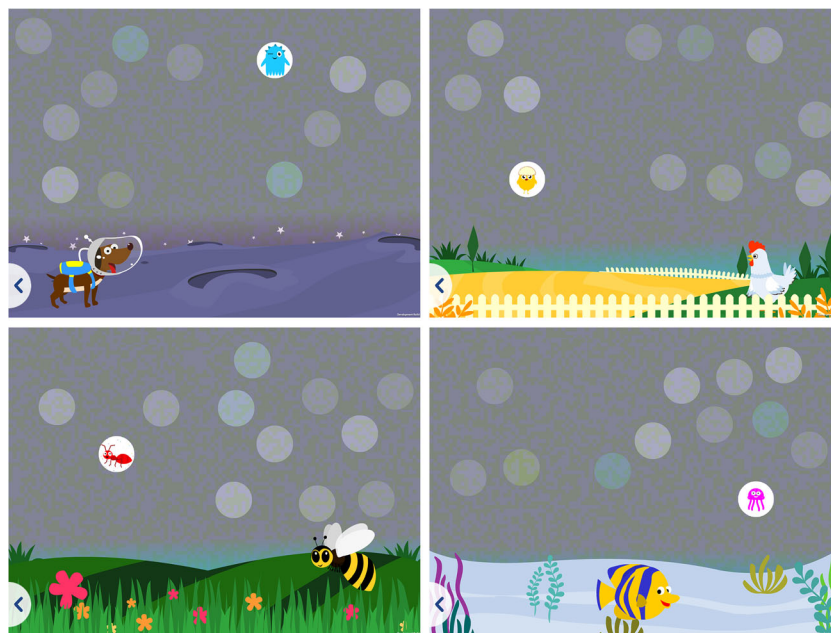


**Fig. 2** Examples of the various scenes showing an animation revealed when a target stimulus is tapped
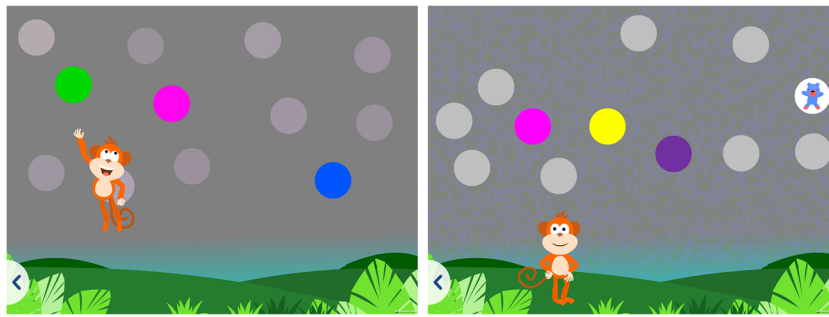
**Fig. 3** (Left) A demonstration tutorial where the monkey character demonstrates the rules of the game by tapping the highly saturated practice targets. **(Right)** A practice trial example where the cartoon animal animation is revealed from behind the colored practice target when the target is successfully tapped

discouragement when they are unable to see the other targets (see Fig. 4 for an example trial simulated for a CVD observer). *ColourSpot* has two sequential sets of three staircases, with the first and then second set ending after the participant has reached 35 trials on each of the protan, deutan and tritan staircases. Once a staircase for a target type has reached 35 trials, the target type is no longer shown but is replaced by a distractor. Each trial remains on the screen until a target or distractor is tapped, and the locations of the targets and distractors are varied randomly on each trial. Children complete the test at their own pace, and at a moderate speed it can be completed in under 5 minutes. Thresholds are computed by fitting psychometric functions (see Results).

## Identifying and validating *ColourSpot*'s classification criteria for CVD

### Participants

A total of 772 boys aged 4–7 years (mean 6.11 years, SD 0.90) were recruited from 27 primary schools in Sussex and London, UK. Each school provided information sheets and consent forms to parents inviting their children to participate

in the research. Only participants with a completed and signed parent/guardian consent form were permitted to participate in the study. Participants were divided into two independent cohorts. The discovery cohort (N = 236, mean 6.17 years, SD 0.83) was used to identify the optimal criteria for assigning CVD versus normal color vision based on the data returned by *ColourSpot*. The validation cohort (N = 536, mean 6.09 years, SD 1.04) was used to test *ColourSpot*'s performance as a diagnostic test for CVD, applying the classification criteria identified using the discovery cohort's data. We chose the sample size for the validation cohort using a custom bootstrap method which indicated that with a sample of 120 control children and 40 children with CVD, we would have 80% power to identify a 99% sensitivity within an 8% range and a 94% specificity within a 12% range. Based on a prevalence of 8%, a sample of 536 boys would be expected to include 43 boys with CVD. The values for sensitivity and specificity are based on those of the Ishihara test for adults since we were aiming to produce a similarly performing test for children. Table 2 provides a breakdown of the numbers of children of each age in the two cohorts.

The study adhered to the World Medical Association's Declaration of Helsinki (2013), with the exception that it was not pre-registered. Ethical approval
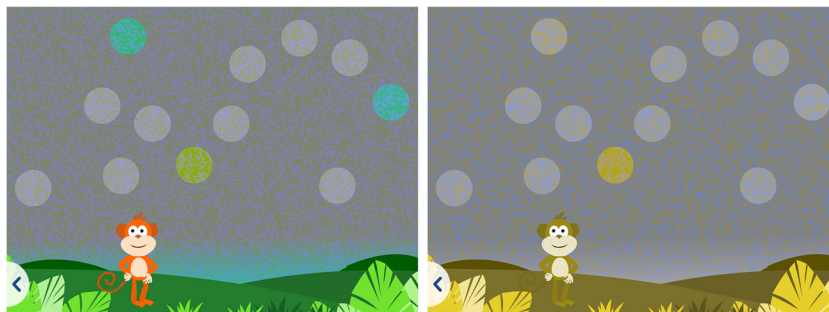


**Fig. 4** A simulation of how a deuteranope may perceive the game (right) compared to a normal trichromat (left). Please note that the simulation is for demonstration purposes only and perception may vary amongst CVD observers. The simulation was produced using Vischeck (Brettel et al., 1997; Lillo et al., 2014)

was granted by the University of Sussex Science & Technology Cross-Schools Research Ethics Committee (ER/TT283/2) and by the European Research Council Executive Agency.

### Color vision tests

All children were administered the Ishihara Unlettered and the Neitz Test. The Ishihara Unlettered (Ishihara & Ishihara, 2016) contains a total of eight plates: three are example plates and five are test plates, including two plates that involve identifying a geometric shape and three plates that involve curve tracing. The Neitz Test is a multiple-choice pen and paper task that has one example plate and eight test plates. Each plate contains the outline of a geometric shape presented against a background of grey dots.

### Procedure

Each participant was screened for CVD using the Ishihara Unlettered and, in order for us to compare *ColourSpot*'s results with those of another test designed for mass-screening, they also completed the Neitz Test. Any participant who made an error or traced irregularly on three or more plates of the Ishihara Unlettered was assigned to a "CVD" group. Any participant who made one or two errors or traced irregularly on the Ishihara Unlettered was assigned to an "inconclusive" group (see Fig. S1 in the Supplementary Information (SI) for the distribution of errors on the Ishihara Unlettered). Any child in the discovery cohort who generated inconclusive results on the Ishihara Unlettered was retested at a later date if possible ($n = 6$), and their group was reassigned if their results were conclusive at retest ($n = 4$ were reassigned to the control group). Children assigned to the inconclusive group in the validation cohort were not retested. A randomly selected subset of children who made no errors on the Ishihara Unlettered was assigned to a "control" group.

The children identified as CVD or inconclusive on the Ishihara Unlettered and the control sample were all tested on *ColourSpot*. Three iPad models (iPad Air 2, 2014; iPad Pro 9.7," 2016; and iPad (5th generation), 2017) were used during

**Table 2** The numbers of participants of each age in the discovery cohort and in the validation cohort

| Age (years) | Discovery Cohort | Validation Cohort |
| --- | --- | --- |
| 4 | 17 | 91 |
| 5 | 95 | 165 |
| 6 | 75 | 162 |
| 7 | 49 | 118 |
| **Total** | **236** | **536** |

testing (also see "Calibration" in the SI). To achieve sufficient trials to measure protan, deutan and tritan discrimination thresholds, participants had to complete a minimum of 40 trials. In the discovery cohort, one participant in the inconclusive group was unable to undertake *ColourSpot* due to school time restraints. This participant was unable to be retested at a later date as they had left the school. In the validation cohort, one participant in the CVD group who had additional educational needs did not want to play *ColourSpot*. Three participants (1 inconclusive, 2 controls) were excluded from the validation cohort as they had completed an insufficient number of trials. Owing to our selection criteria, the administrator of *ColourSpot* was aware of each child's color vision diagnosis by the Ishihara Unlettered when the child was playing *ColourSpot*. However, *ColourSpot* is self-administered by the child with no interference by the administrator once it has started. Therefore, it would be impossible for the administrator to inadvertently bias the child's performance selectively on one color confusion axis in a way that would influence the CVD diagnosis.

## Results

Data from the two staircases for each of the protan, deutan and tritan targets were combined and fit using non-parametric psychometric functions using local-linear fitting via the *modelfree* software package (Żychaluk & Foster, 2009). Figure 5 shows examples of individual psychometric functions from participants in the CVD, inconclusive and control groups. The figure shows that participants in the control group performed similarly for the protan, deutan and tritan targets, whereas participants categorized as CVD by the Ishihara Unlettered had higher thresholds for the protan and deutan targets compared to the tritan targets.

To find the optimal parameters of the model-free estimation of the psychometric function to achieve the best classification of CVD versus normal color vision (according to the Ishihara Unlettered) for the discovery cohort, we made a systematic search of parameter space. The optimal parameters were determined as having a fixed bandwidth of 0.70 for the model-free local linear fit, and a performance level of 0.21 (21%) on which to base threshold estimates (see Fig. S2 in the SI). For each participant and pair of fit parameters we found the minimum (most indicative of protan or deutan performance deficit) of the tritan:protan or tritan:deutan threshold ratio. Then we found the pair of fit parameters which maximized the distance between the CVD participant (according to the Ishihara Unlettered) with the largest ratio and the control participant with the smallest ratio. The model-free estimation was then implemented with the linear algebra library Armadillo (Sanderson & Curtin, 2016, 2018) to be made compatible with Apple software.
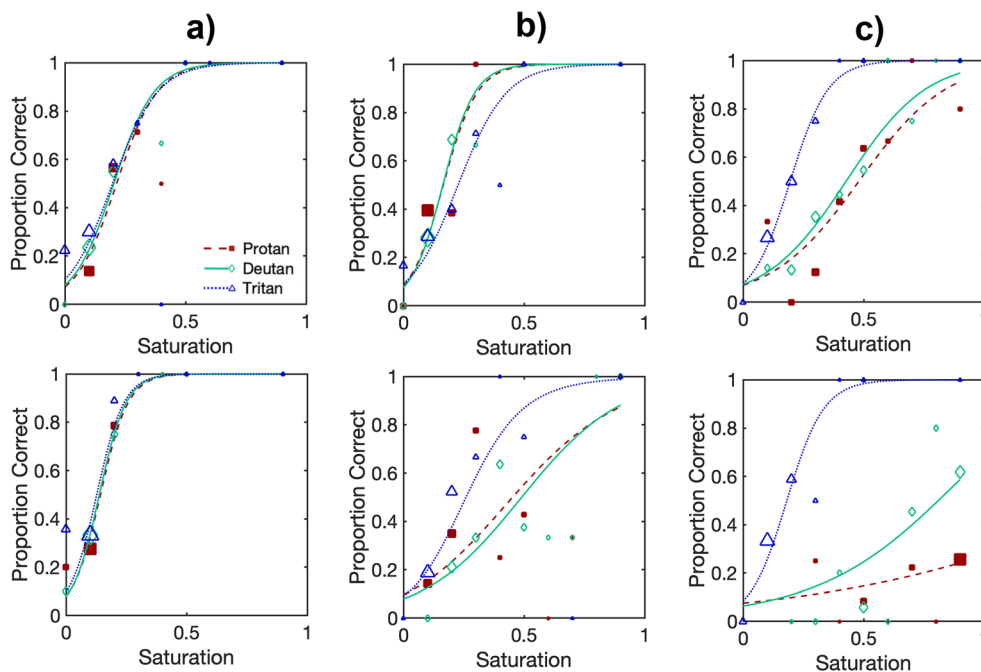
**Fig. 5** Examples of individual psychometric functions for six participants in the Control **(a)**, Inconclusive **(b)** and CVD **(c)** groups fit to their data from *ColourSpot*. Each participant has three psychometric functions representing their performance for detecting protan (red squares, dashed line), deutan (green diamonds, solid line) and tritan (blue triangles, dotted line) targets as a function of *ColourSpot*'s stimulus saturation. The size of the data points is proportional to an arbitrary power of the number of trials of each saturation. *ColourSpot*'s units are relative to the maximum possible saturation of a target on each confusion axis, but the greatest saturation tested on each axis was 0.9 times the maximum. In the chromaticity diagram (MacLeod & Boynton, 1979), the maximum available in gamut saturation (1.0) for protan targets had chromaticity coordinates L/(L+M) = 0.6160, S/(L+M) = 0.0186; the maximum in gamut saturation for deutan targets had chromaticity coordinates L/(L+M) = 0.6160, S/(L+M) = 0.0157; and the maximum in gamut saturation for tritan targets had chromaticity coordinates L/(L+M) = 0.6553, S/(L+M) = 0.0033

## Ability of *ColourSpot* to diagnose CVD
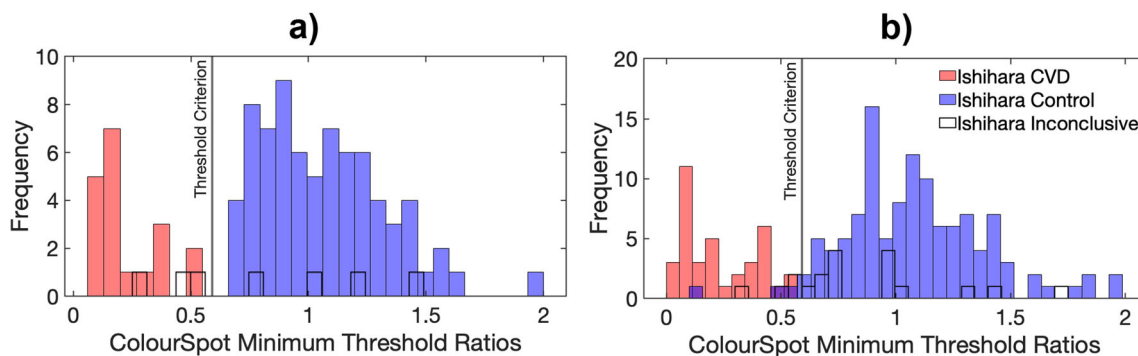
### Discovery cohort

Existing diagnostic tests for CVD are typically based on a criterion score of raw protan and deutan thresholds between individuals (Barbur & Rodriguez-Carmona, 2015; Jurasevska et al., 2014; Mollon et al., 1991; Rabin et al., 2011). We found this to be a suboptimal metric on which to base classification for young children, because the raw thresholds along a given confusion axes do not account for individual differences in non-visual factors such as attention, motivation and engagement that influence task performance (Bosten et al., 2017; Dain & Ling, 2009; Ling & Dain, 2018; see Fig. S3 in the SI for a histogram of our raw protan and deutan thresholds; summary statistics for threshold ratios are provided in Table 4). Instead of raw thresholds, we used as a performance metric the minimum value between the ratio of tritan:protan or tritan:deutan thresholds. We propose this metric is effective at factoring out the influence of non-visual factors on task performance as they would affect thresholds along all three confusion axes equally. Given that congenital tritan deficiencies are extremely rare (Wright, 1952), the risk that a high tritan threshold would cancel high protan and deutan thresholds to create a high ratio in the presence of protan or deutan CVD is very low.

Figure 6a shows a histogram of the threshold ratios for the discovery cohort. There is a clear separation using this metric between groups who were classified either as CVD or as having normal color vision (Control) by performance on the Ishihara Unlettered. Results from the discovery cohort were used to decide on 0.59 as a criterion threshold ratio to define the boundary between a diagnosis of CVD and normal color vision. This was defined as halfway between the maximum threshold ratio for any participant in the CVD group and the minimum threshold ratio for any participant in the control group.

Using the Ishihara Unlettered as a pseudo gold standard test, *ColourSpot* has a nominal sensitivity of 1.00 and a specificity of 1.00 for the discovery cohort (see Table S2 for calculations). Results from *ColourSpot* suggest that the inconclusive group likely includes a combination of participants who have either mild CVD or normal color vision. Figure 6a shows that within the inconclusive group, four out of seven participants perform similarly on *ColourSpot* as participants in the control group (and thus likely have normal color vision).

Fig. 6 Histograms showing, for participants in the discovery cohort (a) and the validation cohort (b), *ColourSpot*'s minimum threshold ratios of the tritan:protan or tritan:deutan thresholds for participants categorized as "CVD," "Inconclusive" and "Control" by the Ishihara Unlettered. The solid vertical line is the criterion threshold ratio used to define the boundary between CVD and normal color vision

Three inconclusive participants' threshold ratios lie within the distribution of threshold ratios for the CVD group (and thus likely have CVD).

## Validation cohort

One hundred and seventy-four participants in the validation cohort played *ColourSpot* (Table 3). The optimal fit parameters and the criterion threshold ratio for defining the boundary between CVD and normal color vision from the results for the discovery cohort were applied without modification to the validation cohort (i.e., the classification method and criteria were independently applied to a new sample). Summary statistics for group threshold ratios for the validation cohort are provided in Table 4.

Again using the Ishihara Unlettered as a pseudo gold standard test, *ColourSpot* achieved a sensitivity of 1.00 and a specificity of 0.97 (see Table S3) for the independent validation cohort. If the criterion threshold ratio were applied to the performances of the inconclusive participants, 15 of the 19 inconclusive participants would be classified as having

normal color vision, and one as having CVD. Three inconclusive participants were near the boundary between the normal and CVD groups and cannot be classified with confidence by either test. In the control group, three participants performed below the threshold criterion. Two of these were near the criterion threshold ratio and cannot be classified with confidence, but one participant in the control group performed similarly on *ColourSpot* to participants in the CVD group, and was likely given a false negative result by the Ishihara Unlettered (see Fig. S4 in the SI and Fig. 6b).

## Neitz Test of Color Vision

A summary of errors on the Neitz Test can be found in Fig. S5 and Fig. S6 of the SI. Briefly, we found that over 50% of boys made errors consistent with a CVD diagnosis. This implies, in agreement with other studies (Tekavčič Pompe, 2020; Tekavčič Pompe & Stirn Kranjc, 2012), that the Neitz Test does not provide an accurate diagnosis of CVD in young children.

## Discussion

Using gamification and psychophysical methods, *ColourSpot* gathers performance data that can be used to classify CVD similarly to or better than the Ishihara Unlettered. *ColourSpot* has many advantages relative to other pediatric tests for CVD and addresses the limitations of these tests. Firstly, our tritan control stimuli and associated "threshold ratio" measure allow us to distinguish visual from non-visual influences on task performance: If relatively poor performance is specific to the protan or deutan confusion lines, CVD is indicated, but if it applies to all three confusion lines, it is likely to be a result of non-visual factors like attention and task engagement. Secondly, gamification with a fun, intuitive and professionally animated interface helps children to maintain engagement and complete the test: 99% of children were

Table 3 Numbers of participants who played *ColourSpot* in the discovery and validation cohorts

| Ishihara Unlettered groups | Discovery cohort | Validation cohort |
|---|---|---|
| CVD | 19 | 37 |
| Inconclusive | 7 | 19 |
| Control | 74 | 118 |
| Exclusions | 0 | 3 |
| Untested | 136 | 362 |
| **Total** | **236** | **536** |

Participants categorized as "CVD" made 3 or more errors on the Ishihara Unlettered. Participants categorized as "Inconclusive" made 1 or 2 errors on the Ishihara Unlettered. Participants categorized as "Control" were randomly selected from children who made no errors on the Ishihara Unlettered. Participants categorized as "Untested" are children who were categorized as having normal color vision by the Ishihara Unlettered but were not tested on *ColourSpot*

**Table 4**  The mean, standard error and 95% confidence intervals of the minimum tritan:protan or tritan:deutan threshold ratios in the discovery and validation cohorts

| Cohort | Ishihara diagnosis | N | Mean | SE | 95% Confidence intervals | |
|---|---|---|---|---|---|---|
| | | | | | Lower bound | Upper bound |
| Discovery | CVD | 19 | 0.23 | 0.03 | 0.16 | 0.29 |
| | Control | 74 | 1.07 | 0.03 | 1.00 | 1.13 |
| | Inconclusive | 7 | 0.83 | 0.16 | 0.43 | 1.23 |
| Validation | CVD | 37 | 0.23 | 0.03 | 0.18 | 0.29 |
| | Control | 117 | 1.08 | 0.03 | 1.03 | 1.14 |
| | Inconclusive | 20 | 0.86 | 0.08 | 0.69 | 1.03 |

able to complete *ColourSpot*, including 4-year-olds and children with additional educational needs. Thirdly, *ColourSpot* is accessible to anyone with an iPad, allowing it to be administered in any setting. Once downloaded, *ColourSpot* automatically provides the relevant calibration file for each iPad model. It is also self-administered, not requiring a trained administrator. This makes *ColourSpot* highly accessible around the world, either for home use with parents, at school with teachers, in the lab with researchers, or in the clinic with optometrists.

The technical features of *ColourSpot* distinguish it from other digitized tests for CVD. The combination of gamification and an adaptive staircase procedure is a psychophysical method well-suited to children. Particularly for children with CVD, the motivation offered by the game's interface and the fact that stimulus contrast is adapted to current task performance both serve to maintain task engagement by reducing discouragement when they are unable to see targets. The addition of luminance and tritan noise allows us to guard against false negatives that might otherwise be caused by individual differences in color vision, particularly amongst individuals with CVD where cone spectral sensitivities vary significantly from the norm (Judd, 1945; Regan et al., 1994). *ColourSpot*'s tritan control stimuli allow our threshold ratio measure, which confers better classification accuracy for CVD versus normal color vision than using raw protan and deutan thresholds. Automatic application of model-specific color calibration allows mass testing and remote testing in different environments, which is an advantage against other digitized tests for CVD that are calibrated only for one particular device (Barbur & Rodriguez-Carmona, 2015; Rabin et al., 2011).

Our validation of *ColourSpot* using two independent cohorts provides statistical strength and gives confidence in the accuracy of classification. The classification algorithm optimized for the discovery cohort was applied to the validation cohort without modification, where it demonstrated similarly strong performance for classifying CVD versus normal color vision relative to the Ishihara Unlettered. For the validation cohort, the sensitivity (again against the Ishihara Unlettered) was 1.00 and the specificity 0.97. We chose the Ishihara

Unlettered pragmatically as a pseudo gold standard because it is not possible to conduct the anomaloscope (the gold standard for adults) on our target age group, owing to its complex task demands. However, the sensitivity and specificity of the Ishihara Unlettered is itself imperfect, evidenced by the fact that it diagnosed 26 children as "inconclusive" in the current study. Using the Ishihara Unlettered as a pseudo gold standard in the current study may therefore lead us to underestimate the sensitivity and specificity of *ColourSpot* compared to using a fully accurate gold standard if diagnostic errors by the Ishihara Unlettered and *ColourSpot* are independent. Alternatively (and less likely), this may lead us to overestimate sensitivity and specificity for *ColourSpot* if diagnostic errors by *ColourSpot* are correlated with those of the Ishihara Unlettered (for example, if a certain type of mild CVD observer is missed by both tests). It is unclear from the Ishihara Unlettered test alone whether the 26 participants that scored inconclusively (1 or 2 errors) on the Ishihara Unlettered were anomalous trichromats or made errors on that test for reasons other than CVD (e.g., lapses in task engagement). However, four of the initially inconclusive participants in the discovery cohort were retested on the Ishihara Unlettered again within 2 months and made no errors, suggesting that they constituted false positive diagnoses by the Ishihara Unlettered. In contrast, *ColourSpot* provides a more secure assignment of color vision status for at least some of the inconclusive participants, suggesting that it may be more accurate at diagnosing CVD than the Ishihara Unlettered. In the validation cohort, one participant was diagnosed as normal by the Ishihara Unlettered but was classified as CVD by *ColourSpot*. Given the position of that participant's threshold ratio near the centre of the CVD distribution (see Fig. 6b and Fig. S6 in the SI for the psychometric function of this participant), we believe that this is likely a false negative diagnosis by the Ishihara Unlettered, but a genetic test would be the only definitive way to confirm the diagnosis at the current age of this participant. As the protan and deutan confusion lines are so close to one another in color space, distinguishing protan CVD from deutan CVD is challenging. It is not possible to estimate *ColourSpot*'s

accuracy for this classification using our current data, as our gold standard test (Ishihara Unlettered) does not accurately distinguish protan CVD and deutan CVD. We plan to explore *ColourSpot*'s ability to classify protan and deutan CVD types by testing it in adults and older children against the anomaloscope.

We aim that accurate diagnosis of CVD using *ColourSpot* from 4 years of age at the start of education will mitigate some of the negative impact of CVD on children's education and well-being (Grassivaro Gallo et al., 1998, 2002; Mehta et al., 2018; Suero et al., 2005; Thomas et al., 2018; Thuline, 1964). The *ColourSpot* app provides advice sheets on CVD, designed by Colour Blind Awareness, a non-profit organization to raise awareness of CVD (http://colourblindawareness.org), which aims to help parents and teachers adopt strategies to mitigate some of the potential negative effects of CVD on children. Once generally released, *ColourSpot* can be used by parents, teachers and optometrists, and we hope that it will enable mass screening for CVD in young children due to its ease of use, psychophysical rigor and diagnostic accuracy.

*ColourSpot* is available to download for research purposes. The source code can be downloaded at https://osf.io/v5p2y/ and/or a fully-compiled version can be made available by contacting the senior authors. As well as the test itself, we also make available for research purposes the interface, animations and general methods which could be applied to measure other visual abilities in children.

## References

Abramov, I., Hainline, L., Turkel, J., Lemerise, E., Smith, H., Gordon, J., & Petry, S. (1984). Rocket-ship psychophysics. Assessing visual functioning in young children. *Investigative Ophthalmology and Visual Science*, 25(11), 1307–1315.

Atowa, U., Wajuihian, S., & Hansraj, R. (2019). A review of paediatric vision screening protocols and guidelines. *International Journal of Ophthalmology*, 12(7), 1194–1201. https://doi.org/10.18240/ijo.2019.07.22

Azizoğlu, S., Crewther, S., Şerefhan, F., Barutchu, A., Göker, S., & Junghans, B. (2017). Evidence for the need for vision screening of school children in Turkey. *BMC Ophthalmology*, 17(1), 230. https://doi.org/10.1186/s12886-017-0618-9

Barbur, J., & Rodriguez-Carmona, M. (2015). Color vision changes in normal aging. In A. J. Elliott, M. D. Fairchild, & A. Franklin (Eds.), *Handbook of Color Psychology* (pp. 180–196). Cambridge University Press. https://doi.org/10.1017/CBO9781107337930.009

Barry, J., Mollan, S., Burdon, M., Jenkins, M., & Denniston, A. (2017). Development and validation of a questionnaire assessing the quality of life impact of Colour Blindness (CBQoL). *BMC Ophthalmology*, 17(1), 1–7. https://doi.org/10.1186/s12886-017-0579-z

Birch, J. (2001). *Diagnosis of defective colour vision*. Butterworth-Heinemann Ltd.

Birch, J. (2012). Worldwide prevalence of red-green color deficiency. *Journal of the Optical Society of America A*, 29(3), 313–320. https://doi.org/10.1364/josaa.29.000313

Birch, J., & McKeever, L. (1993). Survey of the accuracy of new pseudoisochromatic plates. *Ophthalmic & Physiological Optics : The Journal of the British College of Ophthalmic Opticians (Optometrists)*, 13(1), 35–40. https://doi.org/10.1111/j.1475-1313.1993.tb00423.x

Birch, J., & Platts, C. (1993). Colour vision screening in children: an evaluation of three pseudoisochromatic tests. *Ophthalmic and Physiological Optics*, 13(4), 344–349. https://doi.org/10.1111/j.1475-1313.1993.tb00489.x

Block, S., Lee, D., Hoeppner, J., & Birr, A. (2004). Comparison of the Neitz Test of Color Vision to the Ishihara Color Vision tests and the anomaloscopic classification. *Investigative Ophthalmology & Visual Science*, 45(13), 1384.

Bodduluri, L., Boon, M., & Dain, S. (2017a). Evaluation of tablet computers for visual function assessment. *Behavior Research Methods*, 49(2), 548–558. https://doi.org/10.3758/s13428-016-0725-1

Bodduluri, L., Boon, M., Ryan, M., & Dain, S. (2017b). Impact of Gamification of Vision Tests on the User Experience. *Games for Health Journal*, 6(4), 229–236. https://doi.org/10.1089/g4h.2016.0100

Bosten, J. (2019). The known unknowns of anomalous trichromacy. *Current Opinion in Behavioral Sciences*, 30, 228–237. https://doi.org/10.1016/j.cobeha.2019.10.015

Bosten, J., Mollon, J., Peterzell, D., & Webster, M. (2017). Individual differences as a window into the structure and function of the visual system. *Vision Research*, 141, 1–3. https://doi.org/10.1016/j.visres.2017.11.003

Brettel, H., Viénot, F., & Mollon, J. D. (1997). Computerized simulation of color appearance for dichromats. *Journal of the Optical Society of America A*, 14(10), 2647–2655. https://doi.org/10.1364/josaa.14.002647

Chan, X., Goh, S., & Tan, N. (2014). Subjects with colour vision deficiency in the community: What do primary care physicians need to know? *Asia Pacific Family Medicine*, 13(1), 1–10. https://doi.org/10.1186/s12930-014-0010-3

Ciner, E., Dobson, V., Schmidt, P., Allen, D., Cyert, L., Maguire, M., Moore, B., Orel-Bixler, D., & Schultz, J. (1999). A survey of vision screening policy of preschool children in the United States. *Survey*

*of Ophthalmology*, *43*(5), 445–457. https://doi.org/10.1016/s0039-6257(99)00021-1

Cole, B. (2004). The handicap of abnormal colour vision. *Clinical and Experimental Optometry*, *87*(4-5), 258–275. https://doi.org/10.1111/j.1444-0938.2004.tb05056.x

Cole, B. (2007). Assessment of inherited colour vision defects in clinical practice. *Clinical and Experimental Optometry*, *90*(3), 157–175. https://doi.org/10.1111/j.1444-0938.2007.00135.x

Cumberland, P., Rahi, J. S., & Peckham, C. S. (2005). Impact of congenital colour vision defects on occupation. *Archives of Disease in Childhood*, *90*(9), 906–908. https://doi.org/10.1136/adc.2004.062067

Dain, S. (2010). Evaluation of "Colour Vision Testing Made Easy". In J. Mollon, J. Pokorny, & K. Knoblauch (Eds.), *Normal and Defective Colour Vision*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198525301.003.0035

Dain, S., & Almerdef, A. (2016). Colorimetric evaluation of iPhone apps for colour vision tests based on the Ishihara test. *Clinical and Experimental Optometry*, *99*(3), 264–273. https://doi.org/10.1111/cxo.12370

Dain, S., & Ling, B. (2009). Cognitive abilities of children on a gray seriation test. *Optometry and Vision Science : Official Publication of the American Academy of Optometry*, *86*(6), E701-707. https://doi.org/10.1097/OPX.0b013e3181a59d46

de Fez, D., Luque, M., García-Domene, M., Caballero, M., & Camps, V. (2018a). Can Applications Designed to Evaluate Visual Function Be Used in Different iPads? *Optometry and Vision Science*, *95*(11), 1054–1063. https://doi.org/10.1097/OPX.0000000000001293

de Fez, D., Luque, M., Garcia-Domene, M., Camps, V., & Pineero, D. (2016). Colorimetric characterization of mobile devices for vision applications. *Optometry and Vision Science*, *93*(1), 85–93. https://doi.org/10.1097/OPX.0000000000000752

de Fez, D., Luque, M., Matea, L., Piñero, D., & Camps, V. (2018b). New iPAD-based test for the detection of color vision deficiencies. *Graefe's Archive for Clinical and Experimental Ophthalmology*, *256*(12), 2349–2360. https://doi.org/10.1007/s00417-018-4154-y

Department of Health. (2009). *Healthy Child Programme: From 5 to 19 years old*.

Diez, M., Luque, M., Capilla, P., Gomez, J., & de Fez, D. (2001). Detection and assessment of color vision anomalies and deficiencies in children. *Journal of Pediatric Ophthalmology and Strabismus*, *38*(4), 191–195. https://doi.org/10.3928/0191-3913-20010701-05

Farnsworth, D. (1943). The Farnsworth-Munsell 100-Hue and Dichotomous Tests for Color Vision. *Journal of the Optical Society of America*, *33*(10), 568–578. https://doi.org/10.1364/josa.33.000568

Fish, A.-L., Alketbi, M., & Baillif, S. (2020). Evaluation of a New Test for the Diagnosis of Congenital Dyschromatopsia in Children: the Color Vision Evaluation Test. *American Journal of Ophthalmology*, *223*, 348–358. https://doi.org/10.1016/j.ajo.2020.11.003

Grassivaro Gallo, P., Oliva, S., Lantieri, P., & Viviani, F. (2002). Colour Blindness in Italian Art High School Students. *Perceptual and Motor Skills*, *95*(3), 830–834. https://doi.org/10.2466/pms.2002.95.3.830

Grassivaro Gallo, P., Panza, M., Viviani, F., & Lantieri, P. (1998). Congenital dyschromatopsia and school achievement. *Perceptual and Motor Skills*, *86*(2), 563–569. https://doi.org/10.2466/pms.1998.86.2.563

Holroyd, E., & Hall, D. (1997). A re-appraisal of screening for colour vision impairments. *Child: Care, Health and Development*, *23*(5), 391–398. https://doi.org/10.1111/j.1365-2214.1997.tb00906.x

Hopkins, S., Sampson, G., Hendicott, P., & Wood, J. (2013). Review of guidelines for children's vision screenings. *Clinical and Experimental Optometry*, *96*(5), 443–449. https://doi.org/10.1111/cxo.12029

Hovis, J., Leat, S., Heffernan, S., & Epp, K. (2002). The validity of the University of Waterloo Colored Dot Test for color vision testing in adults and preschool children. *Optometry and Vision Science*, *79*(4), 241–253. https://doi.org/10.1097/00006324-200204000-00011

Ichikawa, H., Hukami, K., Tanabe, S., & Kawakami, G. (1979). Standard Pseudoisochromatic Plates Part 2. *Clinical and Experimental Optometry*, *62*, 362. https://doi.org/10.1111/j.1444-0938.1979.tb02331.x

Ishihara, S. (1989). *The series of plates designed as a test for colour-deficiency*. Kanehara.

Ishihara, S. (1998). *Ishihara's tests for colour deficiency*. Kanehara.

Ishihara, S. (2002). *Ishihara's tests for colour deficiency*. Kanehara.

Ishihara, S., & Ishihara, M. (1943). *Ishihara's Design Charts for Colour Deficiency of Unlettered Persons*. Isshinkai Foundation.

Ishihara, S., & Ishihara, M. (2016). *Ishihara's design charts for colour deficiency of Unlettered Persons*. Kanehara Trading Inc.

Jadhav, A., Kumar Sg, P., & Kundu, S. (2017). Importance of colour vision testing in school based eye health examination. *Community Eye Health*, *30*(98), S24–S25.

Judd, D. (1945). Standard Response Functions for Protanopic and Deuteranopic Vision. *Journal of the Optical Society of America*, *35*(3), 199–221. https://doi.org/10.1364/JOSA.35.000199

Jurasevska, K., Ozolinsh, M., Fomins, S., Gutmane, A., Zutere, B., Pausus, A., & Karitans, V. (2014). Color-discrimination threshold determination using pseudoisochromatic test plates. *Frontiers in Psychology*, *5*, 1–7. https://doi.org/10.3389/fpsyg.2014.01376

Kovács, I. (2000). Human development of perceptual organization. *Vision Research*, *40*(10), 1301–1310. https://doi.org/10.1016/S0042-6989(00)00055-9

Lee, D., Cotter, S., & French, A. (1997). Evaluation of Kojima-Matsubara color vision test plates: validity in young children. *Optometry and Vision Science : Official Publication of the American Academy of Optometry*, *74*(9), 726–731. https://doi.org/10.1097/00006324-199709000-00020

Lillo, J., Álvaro, L., & Moreira, H. (2014). An experimental method for the assessment of color simulation tools. *Journal of Vision*, *14*(8), 15. https://doi.org/10.1167/14.8.15

Ling, B., & Dain, S. (2018). Development of color vision discrimination during childhood: differences between Blue–Yellow, Red–Green, and achromatic thresholds. *Journal of the Optical Society of America A*, *35*(4), B35–B42. https://doi.org/10.1364/josaa.35.000b35

MacLeod, D., & Boynton, R. (1979). Chromaticity diagram showing cone excitation by stimuli of equal luminance. *Journal of the Optical Society of America*, *69*(8), 1183–1186. https://doi.org/10.1364/JOSA.69.001183

Mäntyjärvi, M., Tenhunen, J., & Partanen, J. (2000). Screening color vision in preschool-aged children. *Color Research & Application*, *26*(S1), S250–S252. https://doi.org/10.1002/1520-6378(2001)26:1+<::AID-COL52>3.0.CO;2-I

Matsubara, H., & Kojima, K. (1957). *Colour vision test plate for the infants*. Handaya.

Mehta, B., Sowden, P. T., & Grandison, A. (2018). Does deuteranomaly place children at a disadvantage in educational settings? In L. W. MacDonald, C. P. Biggam, & G. V. Paramei (Eds.), *Progress in Colour Studies. Cognition, language and beyond.* (pp. 341–355). John Benjamins Publishing Company. https://doi.org/10.1075/z.217.18meh

Mollon, J., Astell, S., & Reffin, J. (1991). A minimalist test of colour vision. *Documenta Ophthalmologica Proceedings Series*, *54*, 59–67. https://doi.org/10.1007/978-94-011-3774-4_8

Neitz, M., & Neitz, J. (2000). Molecular genetics of color vision and color vision defects. *Archives of Ophthalmology*, *118*(5), 691–700. https://doi.org/10.1001/archopht.118.5.691

Neitz, M., & Neitz, J. (2001). A new mass screening test for color-vision deficiencies in children. *Color Research and Application*, *26*(S1),

S239–S249. https://doi.org/10.1002/1520-6378(2001)26:1+<::aid-col51>3.0.co;2-l

Nguyen, L., Do, E., Chia, A., Wang, Y., & Duh, H. (2014a). DoDo game, a color vision deficiency screening test for young children. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2289–2292. https://doi.org/10.1145/2556288.2557334

Nguyen, L., Lu, W., Do, E., Chia, A., & Wang, Y. (2014b). Using digital game as clinical screening test to detect color deficiency in young children. *ACM International Conference Proceeding Series*, 337–340. https://doi.org/10.1145/2593968.2610486

Parry, N. R. A. (2015). Color vision deficiencies. In A. J. Elliot, A. Franklin, & M. D. Fairchild (Eds.), *Handbook of Color Psychology* (pp. 216–242). Cambridge University Press. https://doi.org/10.1017/CBO9781107337930.011

Pease, P. L., & Allen, J. (1988). A new test for screening color vision: Concurrent validity and utility. *American Journal of Optometry and Physiological Optics*, *65*(9), 729–738. https://doi.org/10.1097/00006324-198809000-00007

Rabin, J., Gooch, J., & Ivan, D. (2011). Rapid quantification of color vision: The cone contrast test. *Investigative Ophthalmology and Visual Science*, *52*(2), 816–820. https://doi.org/10.1167/iovs.10-6283

Ramachandran, N., Wilson, G. A., & Wilson, N. (2014). Is screening for congenital colour vision deficiency in school students worthwhile? A review. *Clinical and Experimental Optometry*, *97*(6), 499–506. https://doi.org/10.1111/cxo.12187

Regan, B. C., Reffin, J. P., & Mollon, J. D. (1994). Luminance noise and the rapid determination of discrimination ellipses in colour deficiency. *Vision Research*, *34*(10), 1279–1299. https://doi.org/10.1016/0042-6989(94)90203-8

Richardson, P., Saunders, K., & McClelland, J. (2008). Colour Vision Testing Made Easy; How Low Can You Go. *Optometry in Practice*, *9*, 17–24.

Rinaldi, L. J., Smees, R., Alvarez, J., & Simner, J. (2020). Do the Colors of Educational Number Tools Improve Children's Mathematics and Numerosity? *Child Development*, *91*(4), e799–e813. https://doi.org/10.1111/cdev.13314

Robbins, S. L., Christian, W. K., Hertle, R. W., & Granet, D. B. (2003). Vision testing in the pediatric population. *Ophthalmology Clinics of North America*, *16*(2), 253–267. https://doi.org/10.1016/s0896-1549(03)00007-5

Sanderson, C., & Curtin, R. (2016). Armadillo: a template-based C++ library for linear algebra. *The Journal of Open Source Software*, *1*(2), 26. https://doi.org/10.21105/joss.00026

Sanderson, C., & Curtin, R. (2018). A user-friendly hybrid sparse matrix class in C++. In J. H. Davenport, M. Kauers, G. Labahn, & J. Urban (Eds.), *Mathematical Software- ICMS 2018* (pp. 422–430). Springer. https://doi.org/10.1007/978-3-319-96418-8_50

Scherf, K., Behrmann, M., Kimchi, R., & Luna, B. (2009). Emergence of Global Shape Processing Continues Through Adolescence. *Child Development*, *80*(1), 162–177. https://doi.org/10.1111/j.1467-8624.2008.01252.x

Sharpe, L., Stockman, A., Jägle, H., & Nathans, J. (1999). Opsin genes, cone photopigments, color vision, and color blindness. In K. R. Gegenfurtner & L. T. Sharpe (Eds.), *Color vision: From genes to perception* (pp. 3–51). Cambridge University Press.

Shute, R., & Westall, C. (2000). Use of the Mollon-Reffin minimalist color vision test with young children. *Journal of American Association for Pediatric Ophthalmology and Strabismus*, *4*(6), 366–372. https://doi.org/10.1067/mpa.2000.110335

Steward, J., & Cole, B. (1989). What do color vision defectives say about everyday tasks? *Optometry and Vision Science : Official Publication of the American Academy of Optometry*, *66*(5), 288–295. https://doi.org/10.1097/00006324-198905000-00006

Stewart-Brown, S. L., & Haslum, M. (1988). Screening of vision in school: Could we do better by doing less? *British Medical Journal*, *297*(6656), 1111–1113. https://doi.org/10.1136/bmj.297.6656.1111

Suero, M. I., Pérez, Á. L., Díaz, F., Montanero, M., Pardo, P. J., Gil, J., & Palomino, M. I. (2005). Does Daltonism influence young children's learning? *Learning and Individual Differences*, *15*(2), 89–98. https://doi.org/10.1016/j.lindif.2004.08.002

Tagarelli, A., Piro, A., Tagarelli, G., Bruno Lantieri, P., Risso, D., & Luciano Olivieri, R. (2004). Colour blindness in everyday life and car driving. *Acta Ophthalmologica Scandinavica*, *82*, 436–442. https://doi.org/10.1111/j.1395-3907.2004.00283.x

Tanabe, S., Ichikawa, H., Hukami, K., & Kawakami, G. (1978). New pseudoisochromatic plates for congenital color vision defects. *Japanese Journal of Clinical Ophthalmology*, *32*(3), 479–487.

Taylor, W. O. (1970). Diagnosis of Colour-Vision Defects in Very Young Children. *The Lancet*, *295*(7650), 781–782. https://doi.org/10.1016/s0140-6736(70)91014-7

Tekavčič Pompe, M. (2020). Color vision testing in children. *Color Research and Application*, *45*(5), 775–781. https://doi.org/10.1002/col.22513

Tekavčič Pompe, M., & Stirn Kranjc, B. (2012). Which psychophysical colour vision test to use for screening in 3–9 year olds? *Zdravstveni Vestnik*, *81*, 170–177.

Teller, D., Pereverzeva, M., & Civan, A. (2003). Adult brightness vs. luminance as models of infant photometry: Variability, biasability, and spectral characteristics for the two age groups favor the luminance model. *Journal of Vision*, *3*(5), 2. https://doi.org/10.1167/3.5.2

Thomas, B. A. W. M., Kaur, S., Hairol, M. I., Ahmad, M., & Wee, L. H. (2018). Behavioural and emotional issues among primary school pupils with congenital colour vision deficiency in the Federal Territory of Kuala Lumpur, Malaysia: A case-control study. *F1000Research*, *7*, 1834. https://doi.org/10.12688/f1000research.17006.1

Thuline, H. C. (1964). Color-Vision Defects in American School Children. *JAMA*, *188*(6), 514–518. https://doi.org/10.1001/jama.1964.03060320036008

Torrents, A., Bofill, F., & Cardona, G. (2011). Suitability of school textbooks for 5 to 7 year old children with colour vision deficiencies. *Learning and Individual Differences*, *21*(5), 607–612. https://doi.org/10.1016/j.lindif.2011.07.004

Verriest, G. (1982). Colour vision tests in children. *Documenta Ophthalmologica Proceedings Series*, *33*, 175–178.

Waggoner, T. (1994). *Colour vision testing made easy*. Waggoner Diagnostics.

WHO Programme for the Prevention of Blindness and Deafness. (2003). *Consultation on development of standards for characterization of vision loss and visual functioning : Geneva, 4-5 September 2003*. World Health Organization. https://apps.who.int/iris/handle/10665/68601

World Medical Association. (2013). World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *Jama, 310*(20), 2191–2194. https://doi.org/10.1001/jama.2013.281053

Wright, W. D. (1952). The Characteristics of Tritanopia. *Journal of the Optical Society of America*, *42*(8), 509–521. https://doi.org/10.1364/JOSA.42.000509

Żychaluk, K., & Foster, D. H. (2009). Model-free estimation of the psychometric function. *Attention, Perception, and Psychophysics*, *71*(6), 1414–1425. https://doi.org/10.3758/APP.71.6.1414