



Development and Validation of a Practical Color Vision Screening Method Using Common Materials Color

Angelea Perez

Psychiatry, School of Medicine, University of Florida, Gainesville, FL, USA
Email: perez.angelea@ufl.edu

How to cite this paper: Perez, A. (2026) Development and Validation of a Practical Color Vision Screening Method Using Common Materials Color. *Open Access Library Journal*, 13: e15257.
<https://doi.org/10.4236/oalib.1115257>

Received: March 27, 2026

Accepted: May 10, 2026

Published: May 13, 2026

Copyright © 2026 by author(s) and Open Access Library Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Purpose: Color Vision Deficiency (CVD) can interfere with educational performance and occupational participation; however, efficient school-based screening tools remain limited. This study evaluated four low-cost color vision screening tests against the Hardy-Rand-Rittler (HRR) Pseudoisochromatic Plates for detecting red-green CVD. **Methods:** Forty-nine adults (35 normal, 14 CVD; 18 - 59 years) completed HRR plates, crayon, color-board, paint-chip classification, and a psychophysical red-increment discrimination task. **Results:** Crayon and color-board tests showed poor reliability due to color desaturation, with overlapping error rates between normal and mild CVD groups. Paint-chip classification improved normal performance but failed to separate mild CVD. In contrast, the red-increment test differentiated CVD from normal vision. At the smallest increment, CVD participants performed near chance (0.28 ± 0.21), whereas normal participants remained above chance (0.68 ± 0.29). ROC analysis demonstrated discrimination (AUC = 0.862 (95% CI: 0.74 - 0.98)), improving with a compound classifier (AUC = 0.906 (95% CI: 0.80 - 0.99)). **Conclusions:** Low-cost color materials lack control; red-increment discrimination offers screening.

Subject Areas

Ophthalmology

Keywords

Color Vision Deficiency, Screening, Psychophysics, Red-Green Deficiency, Pseudoisochromatic Plates, Educational Assessment

1. Introduction

Color Vision Deficiency (CVD), most commonly inherited red-green deficiency

[1] [2], reduces chromatic discrimination and can affect classroom learning, where color is frequently used to encode instructional information. Early identification may support timely educational accommodations and career counseling, particularly because many individuals remain undiagnosed until adolescence or adulthood. Although several established clinical tools exist (e.g., pseudoisochromatic plates, arrangement tests, and anomaloscopes), many are optimized for controlled clinical environments and may be less feasible for school-based screening because they require standardized illumination, specialized materials, or trained administration. In addition, some methods are designed primarily to identify moderate-to-severe deficiencies and may be less sensitive to milder anomalous trichromacy, which is common and relevant to education. [3]

Coren and Hakstian [4] proposed practical specifications for useful color vision testing, emphasizing validation against established measures, reliability, feasibility for group administration, applicability across age ranges, brevity, reproducibility without specialized pictorial plates, interpretability of results as “CVD present/absent”, and independence from prior diagnosis. Guided by these criteria, the present study developed and evaluated low-cost, readily accessible screening procedures intended for eventual primary use in K-12 educational settings. As an initial proof of concept, the tests were first evaluated in adults with known normal color vision or known CVD to establish discriminability under controlled conditions before extension to pediatric populations.

Four economized screening approaches were designed: crayon classification, color-board classification, paint-chip classification, and a psychophysical red-increment discrimination test delivered via digital projection. The first three approaches were intended to exploit characteristic hue/saturation confusions in inherited red-green CVD using commonly available materials [5] [6]. The red-increment test was designed to quantify sensitivity limitations at the long-wavelength end of the spectrum [7] [8], using a constant stimuli framework to estimate detection performance across increment levels.

The primary objective of this study was to determine whether any of these low-cost procedures could provide valid screening outcomes when compared with an established clinical reference test (Hardy-Rand-Rittler pseudoisochromatic plates). We hypothesized that the psychophysical red-increment approach would show the strongest discrimination between participants with CVD and those with normal color vision, whereas classification tasks using consumer color materials would be vulnerable to stimulus variability and desaturation effects unless colorimetric control was ensured.

2. Methods

2.1. Study Design and Participants

This validation study recruited 49 adults (aged 18 - 59 years) via word of mouth and recruitment flyers. Nineteen participants were female, and 30 were male. Color vision status was classified using the Hardy-Rand-Rittler (HRR) Standard

Pseudoisochromatic Plates [9] [10], with 35 participants classified as normal trichromats (16 female, 19 male) and 14 as having CVD (2 female, 12 male). Most participants were recruited from Ewing, New Jersey, and tested at The College of New Jersey. All the participants reported good general and ocular health. Written informed consent was obtained in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of the University of Alabama at Birmingham.

Participants with Color Vision Deficiency (CVD) were recruited through flyers and word-of-mouth within the university community. Several volunteers reported a prior diagnosis or suspicion of color vision deficiency and self-selected into the study. Because this represents a convenience sample, it is likely enriched for individuals with known or suspected CVD and may include a broader severity range than expected in an unselected school population. In typical school populations, mild anomalous trichromacy predominates and overall prevalence is approximately 6% - 8% in males and <1% in females; therefore, this sample may overrepresent both prevalence and severity. No formal sample size calculation was performed; the sample was assembled on a convenience basis as an initial proof-of-concept evaluation. The total sample of 49 provides adequate power to characterize overall CVD versus normal discrimination, but the medium ($n = 2$) and strong ($n = 2$) CVD subgroups are too small for precise severity-stratified estimates; confidence intervals for these subgroups would be wide, and their results should be interpreted descriptively rather than inferentially.

2.2. Reference Standard: HRR Pseudoisochromatic Plates

The HRR Standard Pseudoisochromatic Test, 4th edition [9] [10], was used as the reference standard for the classification of CVD status. The participants viewed the plates at approximately 40 cm. Illumination was provided by a 100-W incandescent bulb in an otherwise dark room; plates were illuminated using a clamp light angled toward the booklet (lamp-to-plate distance of approximately 60 cm). To approximate the standard illuminant conditions, the participants wore daylight conversion filters (Gulden C Daylight glasses). Testing and scoring were performed according to the manufacturer's instructions. The participants were asked to identify and localize the symbols on each plate, and their responses were recorded on a standardized datasheet. A time-limited presentation was implemented for screening the plates (3 s). Based on the screening outcomes, the participants completed the appropriate diagnostic/severity plates as specified by the HRR protocol.

2.3. Economized Screening Tests Using Readily Available Materials

Crayon Classification Test

A prototype screening task was developed using 17 crayons selected from a Crayola 64-count pack, similar to the approaches described by Martins *et al.* [11]

and Neitz and Neitz [12]. Twelve crayons were chromatic samples selected by normal trichromats to represent hues that appeared predominantly blue-green/purple, and five were achromatic (black, white, and gray). Crayon wrappers were covered and each crayon was labeled with a number. Participants judged each sample as either achromatic (“black/white/gray”) or chromatic (“not black/white/gray”) and recorded their responses on a two-column answer sheet. Testing was conducted under typical room lighting (fluorescent ceiling lighting and daylight from windows; ambient daylight was not controlled). For analysis, each participant’s responses were coded as correct or incorrect relative to the trichromat-defined key, and individual error rates were calculated as the proportion correct. The mean error rates were computed separately for the HRR-defined normal and CVD groups.

Color-Board Test

To test whether performance differed when color samples were presented as standardized filled regions rather than crayons, a color board version was created using 17 circles (diameter 1.27 cm) colored with the same crayons used in the crayon test. The participants completed the same dichotomous judgment (chromatic vs. achromatic) and used the same answer sheet and scoring approach described above. The ambient lighting conditions were similar to those in the crayon test.

Paint-Chip Classification Test

A paint-chip sorting task was developed using 17 circular samples (diameter 1.27 cm) cut from commercially available retail paint chips (Benjamin Moore). Two normal trichromats selected samples to approximate the hue range of the crayons. The samples were mounted on a foam-poster board (20.3 × 25.4 cm). Participants again judged each sample as chromatic or achromatic and recorded their responses on the same two-column answer sheet. Scoring and error rate calculations were performed as previously described.

Psychophysical Red-Increment Test

Stimulus Generation and Display

A forced-choice red-increment detection task was developed to assess sensitivity to small, long-wavelength increments [8] [13]. Stimuli were generated in MATLAB (MathWorks, Natick, MA) using the Psychophysics Toolbox [14] and imported into Microsoft PowerPoint (Microsoft Corporation, Redmond, WA) for presentation. Each trial presented a horizontal rectangle of grayscale Gaussian noise (750 × 150 pixels; mean gray level = 100 on a 0 - 255 scale; SD = ±18). A red increment signal was added at one of five possible locations labeled A–E (**Figure 1**). The signal was a two-dimensional Gaussian increment in the red channel (half-width at half-height = 23 pixels) and was spatially dithered (every other pixel) to enable lower effective-contrast levels. Five increment levels were tested with red-channel peak values corresponding to +8, +13, +22, +36, and +60 above the background mean (resulting in peak red values of approximately 108, 113, 122, 136, and 160, respectively). The stimulus subtended approximately 2° at the viewing

distance. The stimuli were projected using an Epson BrightLink Pro wall-mounted projector (Epson America, Inc., Long Beach, CA, USA) onto an interactive whiteboard. The participants viewed the stimuli at approximately 6 m.

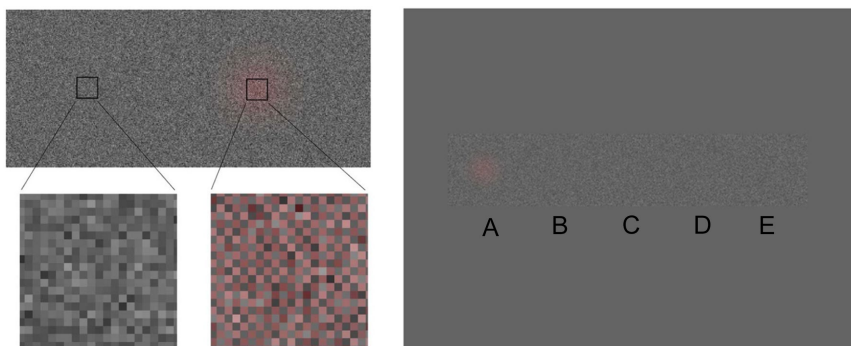


Figure 1. Example stimulus display showing the five response locations (A - E) and a red-increment signal embedded in Gaussian noise.

2.4. Procedure

Trials were presented in blocks of five, with the increment level descending within each block, and the stimulus location was randomized within each block. Each increment level was repeated ten times (50 trials in total). Participants indicated which location contained the red increment (A - E) and were instructed to guess if they were uncertain. The trials were self-paced, and an investigator advanced the slides after each response. Responses were recorded on a 5-column, 50-row, answer sheet.

2.5. Light Measurements and Luminance Estimation

Weber contrast ($\Delta L/L$) was estimated using $\Delta L/L = (L_{\text{increment}} - L_{\text{background}})/L_{\text{background}}$, where $L_{\text{background}} \approx 45 \text{ cd/m}^2$ [2]. Approximate Weber contrasts for the five increment levels (+8, +13, +22, +36, and +60 red-channel units above the mean) were ~ 0.017 , 0.028, 0.050, 0.086, and 0.17, respectively, enabling direct comparison with prior increment-threshold and contrast-sensitivity studies.

The illuminance for each testing condition was measured using a Sekonic L-858D-U meter (Sekonic Corporation, Tokyo, Japan). For the material-based tests, the illuminance at the stimulus surface was approximately 200 - 245 lx; for HRR testing, the illuminance at the plates was approximately 740 lx; illuminant quality substantially affects pseudoisochromatic plate performance. [15] For the red-increment task, the uniform gray background and noise-field luminance were measured at approximately 45 cd/m^2 [16]. Because the Gaussian/dithered increments were not directly measurable with the spot meter at small sizes, the luminance for red-channel values was estimated by measuring large uniform red patches across red-channel settings and fitting a spline to extrapolate the increment values. The mean luminance within the Gaussian center (full width at half maximum) was

computed to estimate the effective increment luminance and Weber contrast. Therefore, absolute increment values should be interpreted as device-specific until a standardized calibration protocol is implemented.

2.6. Statistical Analysis

Analyses were conducted in MATLAB using nonparametric tests due to non-normality (Kruskal-Wallis for group comparisons; Wilcoxon rank-sum for paired distributions, as applicable). Red-increment performance (proportion correct) was modeled as a saturating power function of the increment level. Receiver Operating Characteristic (ROC) analysis was used to evaluate the classification of HRR-defined normal versus CVD groups using red-increment performance at the smallest increment level and a weighted combination of the two smallest increments. [17] Sensitivity and specificity were computed across the criterion thresholds, and the area under the ROC curve (AUC) was calculated using trapezoidal integration. A compound classifier was defined as a weighted linear combination of the two smallest increment performances, with the weight being optimized to maximize the AUC. The AUC values should be interpreted as optimistic estimates pending independent validation.

3. Results

Forty-nine adults (age range, 19 - 59 years) completed the study. Using HRR pseudoisochromatic plates as the reference classification [9], 35 participants were classified as having normal color vision and 14 as having CVD. The CVD group comprised 12 men and two women, consistent with the expected higher prevalence among men [1]. Median age did not differ between groups (CVD: 30 years; normal: 29 years; Wilcoxon rank-sum $p = 0.90$). Among the 14 participants with CVD, the HRR severity classifications were mild ($n = 10$), medium ($n = 2$), and strong ($n = 2$). Red-green subtype classification was frequently “red-green unclassified” in mild cases when HRR errors did not permit clear protan/deutan differentiation (Table 1).

Table 1. Participant demographics and HRR severity classifications (mild, medium, strong) by sex.

Participant ID	Age (years)	Sex	HRR Classification	Severity	Subtype
CVD-01	19	M	CVD	Mild	Red-green unclassified
CVD-02	20	M	CVD	Mild	Red-green unclassified
CVD-03	21	M	CVD	Mild	Red-green unclassified
CVD-04	22	F	CVD	Mild	Red-green unclassified
CVD-05	26	M	CVD	Mild	Red-green unclassified
CVD-06	30	M	CVD	Mild	Red-green unclassified

Continued

CVD-07	32	M	CVD	Mild	Red-green unclassified
CVD-08	35	M	CVD	Mild	Red-green unclassified
CVD-09	38	M	CVD	Mild	Red-green unclassified
CVD-10	42	M	CVD	Mild	Red-green unclassified
CVD-11	45	M	CVD	Medium	Deutan
CVD-12	48	M	CVD	Medium	Protan
CVD-13	52	F	CVD	Strong	Deutan
CVD-14	59	M	CVD	Strong	Protan

Note: All 14 participants with CVD were classified using HRR Standard Pseudoisochromatic Plates, 4th edition. The remaining 35 participants (16 female, 19 male; age range 18 - 58 years; median age 29 years) were classified as having normal color vision. CVD = color vision deficiency; M = male; F = female.

3.1. Performance in Material-Based Color Identification Tasks

Crayon Classification

Normal participants performed near the ceiling: 33/35 (94%) achieved 100% accuracy, with two participants making a single error. In the mild CVD group, 1/10 (10%) of the participants made classification errors. Given the similarly low error rates in normal (6%) and mild CVD (10%) groups, the crayon task did not discriminate mild CVD from normal color vision [11]. Errors among participants with strong CVD differed qualitatively: these participants tended to classify achromatic samples (grayscale crayons) as chromatic, consistent with the findings of Montag and Boynton [18] and Scheibner and Boynton [19].

Color Board (Crayon Marks on White Background)

Compared with direct crayon viewing, the color board increased errors in the normal group (approximately 20% made one or more errors), consistent with reduced saturation when crayons are applied to white paper. Participants with mild CVD demonstrated a similar error rate pattern, indicating limited discriminability between normal and mild CVD using this format. Medium CVD participants made no errors, whereas strong CVD participants showed multiple errors, again primarily misclassifying achromatic samples as chromatic.

3.2. Paint Chips

No normal participant made any errors in the paint-chip task (100% accuracy). Participants with mild CVD made occasional single errors (three participants each made one error), whereas those with medium CVD again made no errors. Participants with strong CVD showed higher error counts, primarily classifying achromatic paint-chip samples as chromatic. Overall, the paint-chip task improved performance among participants with normal vision but did not provide a clear separation of mild CVD from normal color vision, given the low and overlapping error rates (Table 2).

Table 2. Error rates (proportion incorrect) for crayon, color-board, and paint-chip tasks by HRR-defined group.

Group	N	Crayon Test	Color-Board Test	Paint-Chip Test
		Errors (% making ≥ 1 error)	Errors (% making ≥ 1 error)	Errors (% making ≥ 1 error)
Normal	35	2/35 (6%)	7/35 (20%)	0/35 (0%)
Mild CVD	10	1/10 (10%)	2/10 (20%)	3/10 (30%)*
Medium CVD	2	0/2 (0%)	0/2 (0%)	0/2 (0%)
Strong CVD	2	2/2 (100%)	2/2 (100%)	2/2 (100%)

Note: Each test consisted of 17 samples (12 chromatic, 5 achromatic). Participants judged each sample as chromatic versus achromatic. Performance is reported as the number and percentage of participants making one or more classification errors. *Three mild CVD participants each made one error on the paint-chip test. Strong CVD participants primarily misclassified as achromatic.

3.3. Red-Increment Detection Test

Group Performance across Incremental Levels

Across the four largest increments, participants with normal color vision performed with high accuracy (mean proportion correct ≥ 0.93), with no significant differences among those increments (Kruskal-Wallis $p = 0.27$). At the smallest increment (0.75 cd/m [2] equivalent increment), normal participants performed lower but remained well above the 0.20 chance level for a five-alternative forced-choice task (mean \pm SD = 0.68 ± 0.29), which is consistent with normal long-wavelength sensitivity. [7] [20]

Participants classified as having CVD by HRR performed less accurately than normal participants across increment levels. For the three largest increments, the mean performance ranged from approximately 0.83 to 0.90 (Kruskal-Wallis $p = 0.34$). At 1.26 cd/m [2], the mean performance was 0.61 ± 0.26 ($p = 0.001$ vs. larger increments). At the smallest increment (0.75 cd/m [2]), CVD participants performed near chance (0.28 ± 0.21), indicating an inability to detect the lowest increment reliably ($p < 0.001$ vs. larger increments), consistent with the reduced long-wavelength sensitivity in red-green CVD. [8] [20]-[22]

Between-group comparisons demonstrated significantly lower performance in the CVD group relative to the normal group at 0.75, 1.26, 2.24, and 3.89 cd/m [2] (all $p < 0.01$), with no significant difference at 7.69 cd/m [2] ($p = 0.061$) (Figure 2 and Table 3).

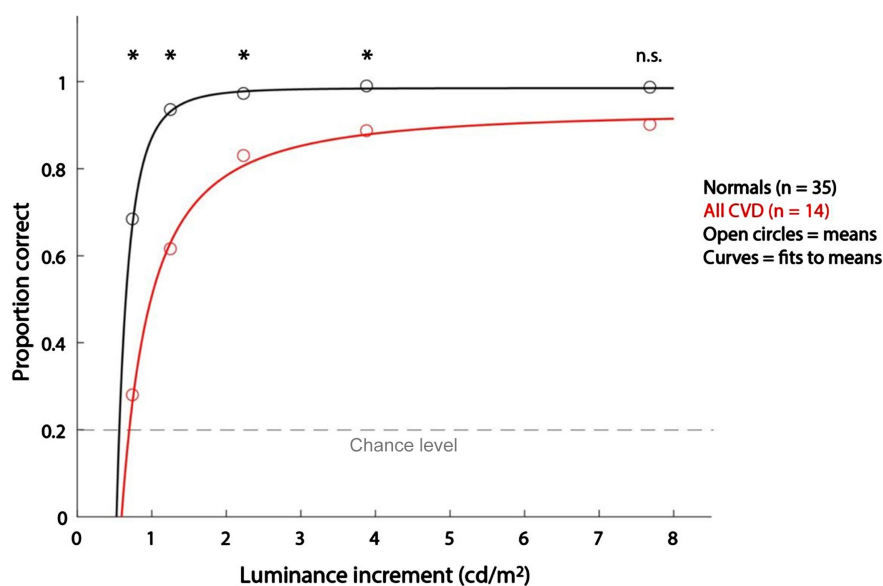
Table 3. Between-group statistical comparisons (Wilcoxon rank-sum) at each increment level.

Group	N	Crayon Test	Color-Board Test	Paint-Chip Test
		Errors (% making ≥ 1 error)	Errors (% making ≥ 1 error)	Errors (% making ≥ 1 error)
Normal	35	2/35 (6%)	7/35 (20%)	0/35 (0%)

Continued

Mild CVD	10	1/10 (10%)	2/10 (20%)	3/10 (30%)*
Medium CVD	2	0/2 (0%)	0/2 (0%)	0/2 (0%)
Strong CVD	2	2/2 (100%)	2/2 (100%)	2/2 (100%)

Note: Each test consisted of 17 samples (12 chromatic, 5 achromatic). Participants judged each sample as chromatic versus achromatic. Performance is reported as the number and percentage of participants making one or more classification errors. *Three mild CVD participants each made one error on the paint-chip test. Strong CVD participants primarily misclassified achromatic samples as chromatic. CVD = color vision deficiency.

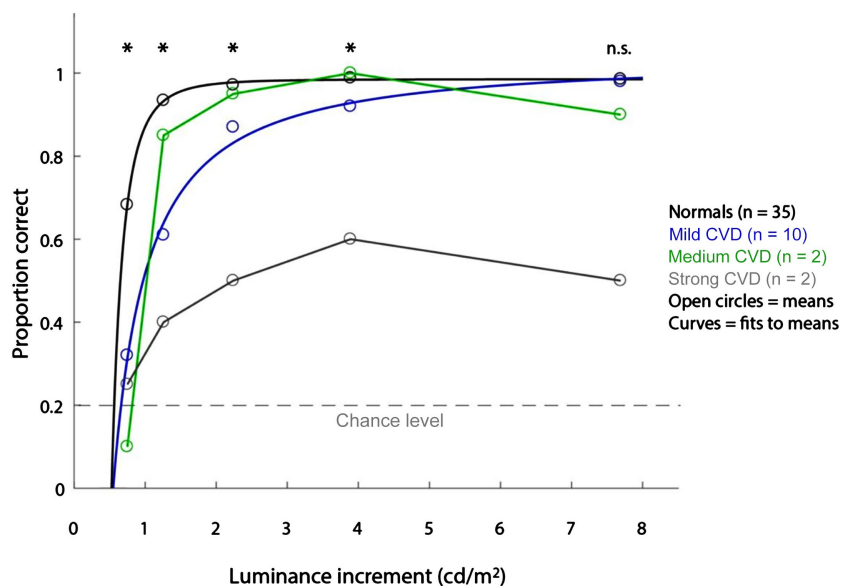


Comparison of normal versus CVD subject performance on the red test, plotted as mean values with curve fits. Significance levels for each increment are as follows: 0.75 cd/m², $p = 7.83 \times 10^{-5}$; 1.26 cd/m², $p = 1.89 \times 10^{-6}$; 2.24 cd/m², $p = 0.0014$; 3.89 cd/m², $p = 0.0041$; 7.69 cd/m², $p = 0.061$ (n.s. = not significant).

Figure 2. Mean red-increment detection performance (proportion correct \pm SD) across five increment levels for normal and CVD groups.

3.4. Severity Subgroup Pattern

When the analysis was restricted to mild CVD ($n = 10$) versus normal participants, mild CVD performance remained significantly lower across the four smaller increment levels (0.75 - 3.89 cd/m [2]; all $p < 0.05$), indicating that the red-increment task distinguished mild CVD from normal performance. The medium and strong CVD subgroups ($n = 2$ each) exhibited lower performance than the other groups but were too small for inferential comparisons; therefore, severity subgroup results should be interpreted descriptively rather than inferentially (Figure 3).

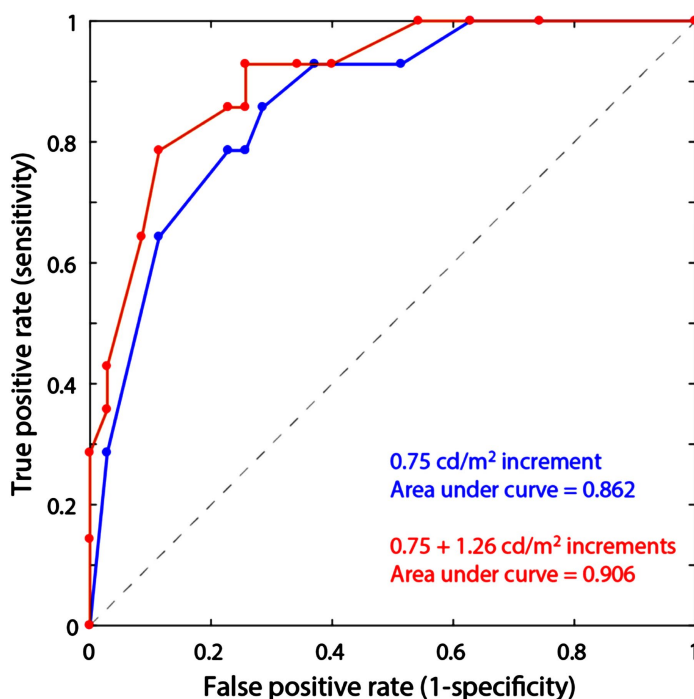


Comparison of normal versus CVD subject performance on the red test, separated by CVD category (mild = blue, medium = green, strong = gray). The proportion of correct trials generally diminished with increasing color deficiency, although the low number for the most affected subjects (drawn with connecting lines) make it impractical to draw any firm conclusions. Significance levels between normals and mild CVDs for each increment are as follows: 0.75 cd/m², $p = 0.0013$; 1.26 cd/m², $p = 1.33 \times 10^{-5}$; 2.24 cd/m², $p = 0.0093$; 3.89 cd/m², $p = 0.015$; 7.69 cd/m², $p = 0.68$ (n.s. = not significant).

Figure 3. Red-increment performance by CVD severity subgroup (mild, medium, strong) versus normal.

3.5. Diagnostic Classification Performance (ROC Analysis)

ROC analysis evaluated the ability of the red-increment task to classify HRR-defined CVD versus normal color vision [17]. Using the percent correct at the smallest increment (0.75 cd/m [2]) as the classifier yielded an AUC of 0.862 (95% CI: 0.74 - 0.98), indicating excellent discrimination. A weighted compound classifier combining the two smallest increments (0.75 and 1.26 cd/m [2]) improved discrimination, yielding an AUC of 0.906 (95% CI: 0.80 - 0.99) at an optimal weight ($w = 0.7$). The moderate width of these confidence intervals reflects the relatively small CVD group ($n = 14$) and should be considered when comparing these estimates with those from larger validation studies. At an example decision criterion of 60% correct for the smallest increment, the sensitivity was 0.857 (12/14) and the specificity was 0.714 (25/35). Criterion optimization indices (Youden and Distance Index) suggested that the optimal cutpoint was near 60% correct (Figure 4, Table 4). Sensitivity and specificity values across all evaluated criterion thresholds for both the single-increment and compound classifiers are summarized in Table 5.



ROC curves for the smallest increment (blue) and a weighted combination of the two smallest increments (red). Dashed line is the lower theoretical limit where test data would yield no classification value (area under curve = 0.5).

Figure 4. ROC curves for single-increment (0.75 cd/m [2]) and compound (0.75 + 1.26 cd/m [2]) classifiers with AUC and 95% CI.

Table 4. Sensitivity, specificity, Youden index, and Distance index at candidate decision criteria.

Classifier	AUC	95% CI	Optimal Criterion	Sensitivity	Specificity	Youden Index	Distance Index
Smallest increment only (0.75 cd/m ²)	0.862	0.746 - 0.978	60% correct	0.857 (12/14)	0.714 (25/35)	0.571	0.371
Compound classifier (w = 0.7)	0.906	0.813 - 0.999	Optimized	0.857 (12/14)	0.829 (29/35)	0.686	0.267

Note: Area Under the Curve (AUC) values indicate diagnostic accuracy for classifying Hardy-Rand-Rittler (HRR)-defined normal versus CVD groups. The compound classifier combined performance at the two smallest increment levels (0.75 and 1.26 cd/m²) using a weighted linear combination with optimized weight (w = 0.7). The optimal criterion for the smallest increment was approximately 60% correct. Youden Index = sensitivity + specificity - 1; Distance Index = $\sqrt{[(1-\text{sensitivity})^2 + (1-\text{specificity})^2]}$. These AUC values should be interpreted as optimistic estimates because classifier weights were optimized on the same dataset; independent validation is required. AUC = Area Under the Curve; CI = Confidence Interval; CVD = Color Vision Deficiency.

Table 5. Sensitivity and specificity across all criterion thresholds for the single-increment (0.75 cd/m²) and compound (0.75 + 1.26 cd/m²) classifiers.

Increment Level	Normal (n = 35)	Mild CVD (n = 10)	Medium CVD (n = 2)	Strong CVD (n = 2)	p-value* (Normal vs Mild)
	Mean ± SD	Mean ± SD	Mean ± SD	Mean ± SD	
0.75 cd/m ²	0.68 ± 0.29	0.32 ± 0.20	0.20 ± 0.14	0.15 ± 0.07	0.001
1.26 cd/m ²	0.93 ± 0.11	0.67 ± 0.24	0.45 ± 0.21	0.40 ± 0.14	0.002
2.24 cd/m ²	0.96 ± 0.08	0.86 ± 0.13	0.75 ± 0.07	0.70 ± 0.14	0.018
3.89 cd/m ²	0.97 ± 0.07	0.90 ± 0.10	0.80 ± 0.14	0.75 ± 0.07	0.042
7.69 cd/m ²	0.98 ± 0.05	0.93 ± 0.08	0.85 ± 0.07	0.80 ± 0.14	0.089

Note: Performance is reported as mean proportion correct ± standard deviation. Mild CVD participants showed significantly lower performance than normal participants across the four smallest increment levels (all $p < 0.05$). Medium and strong CVD subgroups ($n = 2$ each) were too small for inferential comparisons and should be interpreted descriptively. *p-values from Wilcoxon rank-sum tests comparing normal versus mild CVD groups. CVD = Color Vision Deficiency; SD = Standard Deviation.

4. Discussion

This study evaluated four low-cost candidate screening approaches for red-green CVD using HRR pseudoisochromatic plate classification [9]. Material-based dichotomous classification tasks (crayons, crayon color-board marks, and retail paint chips) showed overlapping performance between normal and mild CVD participants, limiting their value as screening tools. In contrast, the psychophysical red-increment detection task demonstrated clear separation between the HRR-defined normal and CVD groups, particularly at the smallest increment, where CVD participants performed near chance and yielded excellent diagnostic discrimination (AUC = 0.862 (95% CI: 0.74 - 0.98) for the smallest increment; AUC = 0.906 (95% CI: 0.80 - 0.99) using a two-increment compound classifier).

4.1. HRR as the Reference Standard and Observed Errors in Normals

HRR pseudoisochromatic plates were used as the reference standard due to their widespread clinical use and ability to categorize defect axis and severity. However, HRR is not a perfect gold standard, and misclassification—particularly among borderline normal and mild CVD cases—is possible. Specifically, a borderline normal observer who is mislabelled as CVD, or a mild CVD observer mislabelled as normal, introduces label noise into the ROC analysis: the classifier is then evaluated against an impure reference, reducing the observed separation between groups. This form of non-differential misclassification would be expected to bias AUC estimates toward 0.5 (the null), attenuating apparent sensitivity and specificity symmetrically. Therefore, the reported AUC values (0.862 and 0.906) likely represent conservative estimates of true diagnostic performance; the actual dis-

crimination of the red-increment task may be somewhat better than these figures suggest. Future studies using anomaloscope-confirmed classification would help quantify this bias.

4.2. Consumer Color Materials Did Not Reliably Discriminate Mild CVD

Both the crayon and color-board tasks were designed as group-administrable classification measures using readily available materials [11] [12]. While normal participants performed well with direct crayons, errors increased substantially with the color-board format (crayon deposition on a white background), and participants with mild CVD showed a similar error pattern. The most plausible explanation is the reduced chromatic saturation and increased reliance on luminance/brightness cues when the pigment is applied thinly or inconsistently to a high-reflectance background. This is consistent with broader evidence that desaturated color tasks tend to increase false-positive rates and reduce discriminability in screening contexts. [5] [6]

The paint-chip task improved performance among normal participants (no errors observed) and produced occasional errors among mild CVD participants; however, the overlap remained too large for screening, and the task remained vulnerable to uncontrolled stimulus properties. Because consumer paint formulations and printing/production processes can vary, and because the intended near-confusion-line sampling was selected by trichromats without spectrophotometric verification [23], the paint-chip approach cannot be assumed to consistently sample the chromatic regions required to maximize CVD detection. [24] [25]

Red-Increment Detection Demonstrated Robust Discrimination

The red-increment task was developed to leverage a well-established feature of red-green CVD: reduced sensitivity in the long-wavelength region under conditions that minimize compensatory cues. [7] [8] [20] [24] In this study, normal participants showed suprathreshold performance even at the smallest increment, whereas HRR-classified CVD participants performed near chance at that same increment, indicating a functional inability to detect the increment and consequent guessing in the forced-choice task. Importantly, the red-increment method remained discriminative even when analyses were restricted to mild CVD participants, which is critical for screening because mild anomalous trichromacy is common and often overlooked by categorical color-identification tasks. [3]

ROC analysis further supports this conclusion: the classification performance was excellent using the smallest increment alone and improved when incorporating information from the second smallest increment. These AUC values should be interpreted as optimistic estimates because the classifier weights were optimized on the same dataset; therefore, independent validation is required.

4.3. Red-Increment Detection Demonstrated Robust Discrimination

The red-increment task was developed to leverage a well-established feature of

red-green CVD: reduced sensitivity in the long-wavelength region under conditions that minimize compensatory cues [7] [8] [20] [24]. In this study, normal participants showed suprathreshold performance even at the smallest increment, whereas HRR-classified CVD participants performed near chance at that same increment, indicating a functional inability to detect the increment and consequent guessing in the forced-choice task. Importantly, the red-increment method remained discriminative even when analyses were restricted to mild CVD participants, which is critical for screening because mild anomalous trichromacy is common and often overlooked by categorical color-identification tasks. [3]

ROC analysis further supports this conclusion: the classification performance was excellent using the smallest increment alone and improved when incorporating information from the second smallest increment. These AUC values should be interpreted as optimistic estimates because the classifier weights were optimized on the same dataset; therefore, independent validation is required.

4.4. Practical Implications for School-Based Screening

From an implementation perspective, the red-increment paradigm has several advantages aligned with school screening needs [4]: it is brief, does not require specialized printed plates, and can be delivered with widely available display or projection technologies supported by validated stimulus software [14]. In contrast, physical sample tests (crayons, marks on paper, and paint chips) may be appealing for logistical reasons but are highly sensitive to stimulus saturation, reflectance, lighting spectrum, sample wear/soiling, and manufacturing variability. If physical materials are used for screening in educational environments, they must be standardized and periodically verified to avoid drift that could compromise accuracy [15] [25].

5. Limitations

This study had several limitations. First, the sample size for the medium- and strong-CVD categories was small, limiting severity-stratified inference; these subgroup results should be interpreted descriptively. Second, the material-based tasks were conducted under typical room lighting with a daylight contribution that was not tightly controlled, which reflects real-world screening conditions but also introduces variability in the color appearance of test stimuli [26]. Third, the red-increment luminance increments were estimated indirectly because of the stimulus spatial properties (Gaussian and dithering) and should therefore be described as device-specific equivalent increments. Finally, HRR was used as the reference standard. Future validation should include anomaloscope-based confirmation of CVD subtype and severity, where feasible. [27] [28]

Methodological Improvements and Future Studies

Future studies should focus on optimizing the red-increment task for group screening and younger age groups [13] [29]. Based on the observed redundancy at higher increments and strong discriminability at the lowest increments, the task

could be shortened substantially by using fewer increment levels (e.g., two or three levels) and fewer trials per level while maintaining a forced-choice structure. For early elementary populations, response formats should be developmentally appropriate (e.g., shape-based response options rather than letter labels). A group-administration workflow is feasible if stimulus timing and response capture are simplified (for example, using classroom response cards or tablet-based entry), although forced-choice performance should remain the primary scoring basis to preserve psychophysical interpretability.

The next critical step is the direct evaluation of the intended population (K-12), including the assessment of feasibility, test-retest reliability, and rate of false positives under school lighting. If the goal is broad deployment, a calibration procedure (or, at minimum, device qualification specifications) will be required to ensure that the red-increment contrast levels are stable across different projection or display technologies and ambient conditions.

6. Conclusion

This study evaluated several low-cost, easily deployable approaches for screening red-green CVD against a widely used clinical reference (HRR pseudoisochromatic plates) [9]. Material-based classification tasks using crayons, crayon marks on paper, and retail paint chips did not provide sufficient discriminative power to reliably separate individuals with normal color vision from those with mild CVD, largely due to stimulus desaturation and a lack of chromatic standardization [5] [6]. In contrast, a psychophysical red-increment detection task delivered via commonly available projection technology demonstrated robust discrimination between HRR-classified normal and CVD participants, including those with mild deficits in HRR [8] [13]. The diagnostic performance was excellent (AUC approaching 0.9) [17], and CVD participants performed near chance at the smallest increment level, while normal participants remained reliably above chance. Although further validation is required in larger and younger populations and across different display technologies and lighting environments, these results support the red-increment paradigm as a promising foundation for brief group-administered screening in educational settings. [4]

Conflicts of Interest

The author declares no conflicts of interest.

References

- [1] Birch, J. (2012) Worldwide Prevalence of Red-Green Color Deficiency. *Journal of the Optical Society of America A*, **29**, 313-320. <https://doi.org/10.1364/josaa.29.000313>
- [2] Neitz, J. and Neitz, M. (2011) The Genetics of Normal and Defective Color Vision. *Vision Research*, **51**, 633-651. <https://doi.org/10.1016/j.visres.2010.12.002>
- [3] Knoblauch, K., Vital-Durand, F. and Barbur, J.L. (2001) Variation of Chromatic Sensitivity across the Life Span. *Vision Research*, **41**, 23-36. [https://doi.org/10.1016/s0042-6989\(00\)00205-4](https://doi.org/10.1016/s0042-6989(00)00205-4)

- [4] Coren, S. and Hakstian, A.R. (1998) Testing Color Discrimination without the Use of Special Stimuli or Technical Equipment. *Behavior Research Methods, Instruments, & Computers*, **30**, 495-503.
- [5] Cole, B.L., Lian, K., Sharpe, K. and Lakkis, C. (2006) Categorical Color Naming of Surface Color Codes by People with Abnormal Color Vision. *Optometry and Vision Science*, **83**, 879-886. <https://doi.org/10.1097/01.opx.0000249974.22205.2a>
- [6] Montag, E.D. (1994) Surface Color Naming in Dichromats. *Vision Research*, **34**, 2137-2151. [https://doi.org/10.1016/0042-6989\(94\)90323-9](https://doi.org/10.1016/0042-6989(94)90323-9)
- [7] Hsia, Y. and Graham, C.H. (1957) Spectral Luminosity Curves for Protanopic, Deutanopic, and Normal Subjects. *Proceedings of the National Academy of Sciences*, **43**, 1011-1019. <https://doi.org/10.1073/pnas.43.11.1011>
- [8] Loop, M.S., Shows, J.F., Mangel, S.C. and Kuyk, T.K. (2003) Colour Thresholds in Dichromats and Normals. *Vision Research*, **43**, 983-992. [https://doi.org/10.1016/s0042-6989\(03\)00074-9](https://doi.org/10.1016/s0042-6989(03)00074-9)
- [9] Hardy, L.H., Rand, G. and Rittler, M.C. (2002) HRR Pseudoisochromatic Plates. 4th Edition, Richmond Products.
- [10] Bailey, J.E., Neitz, M., Tait, D.M. and Neitz, J. (2004) Evaluation of an Updated HRR Color Vision Test. *Visual Neuroscience*, **21**, 431-436. <https://doi.org/10.1017/s0952523804213463>
- [11] Martins, G.M., Bordaberry, M.F., Corrêa, Z.M.S., Mânica, M.B., Costa, J.C., Telichevesky, N., et al. (2001) Color Vision in School-Age Children: Assessment of a New Test. *Jornal de Pediatria*, **77**, 327-30. <https://doi.org/10.2223/jped.242>
- [12] Neitz, J. and Neitz, M. (2000) A New Mass Screening Test for Color-Vision Deficiencies in Children. *Optometry and Vision Science*, **77**, 364-370.
- [13] York, Y.C. and Loop, M.S. (2008) Red Light Increment Threshold as a Measure of Deficient Color Vision. *Optometry and Vision Science*, **85**, 106-111. <https://doi.org/10.1097/opx.0b013e3181622648>
- [14] Brainard, D.H. (1997) The Psychophysics Toolbox. *Spatial Vision*, **10**, 433-436. <https://doi.org/10.1163/156856897x00357>
- [15] Dain, S.J. (1998) Daylight Simulators and Colour Vision Tests. *Ophthalmic and Physiological Optics*, **18**, 540-544. <https://doi.org/10.1046/j.1475-1313.1998.00380.x>
- [16] Koenig, D. and Hofer, H. (2011) The Absolute Threshold of Cone Vision. *Journal of Vision*, **11**, Article No. 21. <https://doi.org/10.1167/11.1.21>
- [17] Thiadens, A.A.H.J., Hoyng, C.B., Polling, J.R., Bernaerts-Biskop, R., van den Born, L.I. and Klaver, C.C.W. (2013) Accuracy of Four Commonly Used Color Vision Tests in the Identification of Cone Disorders. *Ophthalmic Epidemiology*, **20**, 114-122. <https://doi.org/10.3109/09286586.2012.759596>
- [18] Montag, E.D. and Boynton, R.M. (1987) Rod Influence in Dichromatic Surface Color Perception. *Vision Research*, **27**, 2153-2162. [https://doi.org/10.1016/0042-6989\(87\)90129-5](https://doi.org/10.1016/0042-6989(87)90129-5)
- [19] Scheibner, H.M.O. and Boynton, R.M. (1968) Residual Red-Green Discrimination in Dichromats. *Journal of the Optical Society of America*, **58**, 1151-1158. <https://doi.org/10.1364/josa.58.001151>
- [20] Schwartz, S.H. (1994) Spectral Sensitivity of Dichromats: Role of Postreceptor Processes. *Vision Research*, **34**, 2983-2990. [https://doi.org/10.1016/0042-6989\(94\)90270-4](https://doi.org/10.1016/0042-6989(94)90270-4)
- [21] Wandell, B.A. (1987) The Synthesis and Analysis of Color Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **9**, 2-13.

- <https://doi.org/10.1109/tpami.1987.4767868>
- [22] Kaiser, P.K., Lee, B.B., Martin, P.R. and Valberg, A. (1990) The Physiological Basis of the Minimally Distinct Border Demonstrated in the Ganglion Cells of the Macaque Retina. *The Journal of Physiology*, **422**, 153-183. <https://doi.org/10.1113/jphysiol.1990.sp017978>
- [23] Pitt, F.H.G. (1935) Characteristics of Dichromatic Vision. Medical Research Council Special Report Series No. 200. His Majesty's Stationery Office.
- [24] Cole, B.L. (2004) The Handicap of Abnormal Colour Vision. *Clinical and Experimental Optometry*, **87**, 258-275. <https://doi.org/10.1111/j.1444-0938.2004.tb05056.x>
- [25] Dain, S.J. (2004) Colorimetric Analysis of Four Editions of the Hardy-Rand-Rittler Pseudoisochromatic Tests. *Visual Neuroscience*, **21**, 437-443. <https://doi.org/10.1017/s0952523804213475>
- [26] Birch, J. (2001) Diagnosis of Defective Colour Vision. 2nd Edition, Butterworth-Heinemann.
- [27] Thomas, P.B.M. and Mollon, J.D. (2004) Modelling the Rayleigh Match. *Visual Neuroscience*, **21**, 477-482. <https://doi.org/10.1017/s095252380421344x>
- [28] Birch, J. (1997) Clinical Use of the American Optical Company (Hardy, Rand and Rittler) Pseudoisochromatic Plates for Red-Green Colour Deficiency. *Ophthalmic and Physiological Optics*, **17**, 248-254. <https://doi.org/10.1111/j.1475-1313.1997.865.x>
- [29] Vital-Durand, F. and Cottard, A. (1985) Preferential Looking and Early Color Vision. *Vision Research*, **25**, 729-735.