Military Research ColorDx and Printed Color Vision Tests

Ali Almustanyir; Jeffery K. Hovis

PURPOSE: To determine the equivalence of the ColorDx Military Research version (mColorDx) test and three printed pseudoisochromatic tests (HRR, Ishihara, and PIPIC) for color vision testing.

METHODS: Participating in the study were 75 color-normals and 47 subjects with red-green color vision defects. Color vision was classified by an anomaloscope. The HRR (4th edition), Ishihara 38-plate edition, and PIPIC tests are printed color vision tests, whereas mColorDx test figures were displayed on a calibrated computer desktop monitor. All tests were repeated in about 1 wk.

RESULTS: The kappa level of agreement (κ) values with the anomaloscope for screening for each test was 0.96 or greater. The values were statistically identical. Specificity for each test was at least 0.99 and sensitivity was at least 0.95. The repeatability of the screening sections for all tests was very good with κ values greater than 0.95. Deutans tended to miss the tritan screening plates on the HRR and mColorDx tests. The Spearman rank correlation coefficients between the severity of the defect and anomaloscope range was moderate with $r = 0.45$ for the mColorDx and $r = 0.6$ for the HRR. Both the mColorDx and HRR had perfect agreement with the anomaloscope in classifying the defects as either protan or deutan.

CONCLUSION: The validity of the four tests for color vision screening was statistically identical; however, the HRR may be preferred because it had the highest sensitivity of 0.99, a specificity of 1.0, and a reasonable correlation between the severity rating of the defect and the anomaloscope range.

KEYWORDS: color vision test, mColorDx, ColorDx Military Research version, HRR, Ishihara, PIPIC.

Almustanyir A, Hovis JK. *Military Research ColorDX and printed color vision tests*. Aerosp Med Hum Perform. 2015; 86(10):852–859.

The number of color vision tests has increased recently and so it becomes important to know which ones are the most convenient, valid, and reliable. Knowing how to administer a test is not sufficient. One should also have an understanding of the test design in order to have high confidence in interpreting an individual's result.[4,9] The purpose of the color vision test can be classified into three levels:[4]

- To screen for color vision defects. If a given test can accurately and quickly divide subjects into defective color vision and normal color vision, then this test is a good screening test.
- To determine the type and severity of color vision defect.
- To determine whether they have adequate color vision to carry out a specific occupational task. Individuals with congenital color vision deficiencies are at a greater risk for making an error in color judgment and this is the primary reason for color vision testing for occupations where color recognition and discrimination is important.

Color vision deficiencies can be divided into congenital or acquired.[10] The primary features that distinguish congenital from acquired are that the visual system in a congenital defect is otherwise normal except for the loss in color discrimination and the defect remains stable throughout life.[10] In contrast, acquired color vision defects are always due to an underlying disease or disorder and some other aspect of visual function is usually affected by the condition. Furthermore, the acquired color vision defect can regress and progress along with the underlying condition.[10]

From the School of Optometry and Vision Science, University of Waterloo, Waterloo, Ontario, Canada, and the College of Applied Medical Sciences, Optometry Department, King Saud University, Riyadh, Saudi Arabia.

This manuscript was received for review in March 2015. It was accepted for publication in June 2015.

Address correspondence to: Ali Almustanyir, School of Optometry and Vision Science, University of Waterloo, 200 University Avenue West, Waterloo, ON, Canada N2L 3G1; a3almust@uwaterloo.ca.

DOI: 10.3357/AMHP.4326.2015

Congenital color vision defects are the most common color vision deficiency in patients under 60 yr old. After the age of 60 yr, acquired defects become more common.[13] Congenital color vision deficiencies are classified as either red-green or blue-yellow based on which colors they are more likely to confuse. Individuals with red-green defects confuse colors along the red-green color axis (e.g., red, orange, yellow, and green), whereas individuals with blue-yellow defects confuse colors along the blue-yellow color axis (e.g., violet, gray, and yellow-green). Within these two broad categories, the defect can be divided into dichromatic and anomalous trichromatic based on the number of primaries required to match a colored light. Clinically, the distinction between these two categories is often impossible unless the clinician has an anomaloscope and so the severity of the defect is usually graded as mild, moderate, or severe.

Red-green color defectives can be further classified based on whether the M-cone or L-cone photopigment is missing or different from the color-normal population. Deutans have a different or missing M-cone photopigment and protans have a different or missing L-cone photopigment.[5] Congenital blue-yellow defects or tritan defects are very rare. The function of the S-cone photopigment is compromised to varying degrees in these individuals.[5]

Color vision screening tests are usually based on a pseudoisochromatic design where a number or figure is either not seen by a person with a color vision defect or the color-defective person perceives a different number or figure.[2,4,10] Some of these clinical tests also grade the severity and classify the red-green defects as either protan or deutan. Until recently, all the clinical pseudoisochromatic color vision tests were in the printed form. This changed with the increased number of computerized color vision tests now available, such as the Cone Contrast Test and the ColorDx test.

The purpose of this study was to compare three commonly used printed pseudoisochromatic color vision tests with the computerized Military Research (ver. 2011) of the ColorDx (mColorDx) test (Konan Medical, Irvine, CA) with respect to their screening validity, classification validity, and repeatability. The three printed tests are the 4th edition of the Hardy Rand Rittler test (HRR; Richmond Products, Albuquerque, NM), the 38-plate edition of the Ishihara test (Ishihara; Kanehara & Co., Ltd, Tokyo, Japan, 1996), and the 1st edition Pseudoisochromatic plates Ishihara Compatible (PIPIC; T.L. Waggoner, Inc; Gulf Breeze, FL, 2005). The HRR and Ishihara tests are common clinical screening color vision tests.[3,4] The advantages of the HRR over the Ishihara test are that it can screen for both blue-yellow and red-green defects and it uses simple shapes instead of numbers. The PIPIC test is similar in design to the Ishihara test, although it can screen for both red-green and blue-yellow defects. The HRR has 2 blue-yellow and 4 red-green screening plates along with 4 blue-yellow and 10 red-green diagnostic plates. The Ishihara test has 20 red-green screening plates and 4 diagnostic plates. The PIPIC has 13 red-green screening plates, 2 blue-yellow screening plates, and 1 classification plate. To our knowledge, there are no studies that have investigated the validity of this test.

The mColorDx test is a pseudoisochromatic design computer test that presents numbers as the test figures. The desktop version was used in this study. The screening portion of the test presents up to 25 images that screen for red-green defects followed by 12 plates that screen and estimate the severity of any blue-yellow defect. The screening images resemble the Ishihara vanishing and transformation plates. The blue-yellow plates are the vanishing design with colored numbers within a gray background. The figure colors are along the blue-yellow line of confusion with gray and vary in saturation. If the person fails the red-green screening series, then 64 diagnostic plates are administered. All of the red-green diagnostic plates have a gray background with half of the plates having a figure color along the protan line of confusion and the other half having a figure color along the deutan line of confusion. The saturation varies in each series so that the severity of the defect is estimated by the number of saturation levels missed by the subjects. The protan series is presented before the deutan series. The diagnostic and blue-yellow test plates start with the most saturated test figures and progress to the least saturated figures.

METHODS

Through advertisements in the local papers and social media, 75 color normal subjects and 47 subjects with a red-green color vision defect were recruited. The sample size was selected so that the standard error of the estimated kappa coefficient (κ) for agreement between the HRR and Ishihara with the Nagel anomaloscope would be less than or equal to 0.02. Color vision was classified according to the Nagel Anomaloscope using the procedure outlined by Schmidt for neutral adaptation.[12] The color-normal participants were 60% females and 40% males, whereas the color abnormal group was predominantly male (94% males and 6% females). The difference in the proportions of males and females in the two groups was due to X-linked recessive inheritance of red-green color vision deficiencies. The numbers of color-defectives were 11 (23%) deuteranomalous, 23 (49%) deuteranopes, 5 (11%) protanomalous, and 8 (17%) protanopes. The subjects' ages ranged from 16 to 71 yr. The 16 to 71 age range was primarily related to the availability of color-defectives who were willing to participate in the study. The distributions of ages were not normally distributed (Shapiro-Wilk normality test; failed $P < 0.050$) and so the Mann-Whitney rank sum test was performed to determine whether there was a statistically significant difference in the ages of the two groups. The median age was 24 yr for the color-normals and 25 yr for the color-defects. These values were not statistically significant ($P = 0.492$). There were no self-reported vision problems other than a color vision defect or a corrected refractive error. The possibility of a bilateral disorder associated with an acquired color vision defect was reduced further by restricting the subject pool to only those with a visual acuity of at least 6/6 binocularly with or without corrective lenses. Presbyopic subjects wore their current near prescription during the tests.

The printed color vision tests were administered as follows. Subjects viewed these tests from approximately 60 cm. Each test started with the demonstration plate to make sure that the subject understood the test. They were asked to identify either the shape(s) or number(s) on each plate within approximately 5 s, but this was not well controlled. The tests were administered under an Illuminant C fluorescent lamp with an illuminance on the test booklets of 1400 lx.

The mColorDx test was displayed on a LG monitor (Model:W2442PAT) using a PC computer with a Windows 7 Professional operating system. The monitor was calibrated according to test instructions using the Spyder program (4PRO ver. 4.4.5) to a white reference of 6500 K correlated color temperature. The plate was presented within a white background with an average luminance of 102.5 $cd \cdot m^{-2}$ (SD $\pm$ 6.2 $cd \cdot m^{-2}$). The diameter of the color background was 5.1° and the average angular height of the numbers was 3° (SD $\pm$ 0.3°). Each plate was presented for 2 s. After the test figure disappeared, a list of nine numbers appeared on a white screen and the subject selected which of the nine numbers was seen on the plate. The possible choices remained visible until the subject responded. Although there were 25 screening plates, the red-green screening test ended once a total of 5 errors were made and the program switched to the diagnostic series starting with the blue-yellow series.

The order of the tests was the PIPIC, Ishihara, HRR, and the mColorDx test for all subjects. The tests were repeated in the same order after a minimum of 5 d from first session. The experiment was reviewed and approved by the University of Waterloo's Office of Research Ethics. All subjects gave written informed consent before participating.

Pass/Fail Criteria for Screening

The failure criteria for the screening portion of each test were as follows:

HRR. Any error on the red-green or blue-yellow screening figures was a failure of the respective section.

Ishihara. More than 3 errors on the transformation and vanishing plates (i.e., plates 1-17).

PIPIC. More than 2 errors on the red-green screening plates and any error on the blue-yellow plates.

mColorDx. More than 4 errors was a failure on the red-green screening plates and more than 2 errors was a failure on the blue-yellow screening plates.

Answers were scored as correct only if the subject's response matched the score key exactly; otherwise, the response was marked as an error. This criterion was adopted to be consistent with the mColorDx, which does not allow for any interpretation of the response. The response entered was either correct or incorrect. The criteria for the mColorDx and PIPIC were the manufactures' recommended criterion. The HRR criteria were selected to maximize the red-green sensitivity for a single presentation of the screening plates according to Cole et al.'s study.[3] The Ishihara criterion was based on Birch's recommendation for administering only the transformation and vanishing plates.[2] The type of defect for all tests was based on whether a simple majority of errors were on the deutan or protan diagnostic figures. In this study, none of the subjects was asked which of the two diagnostic figures was more visible if both were seen on any of the printed tests. The severity of the color vision defect identified by the HRR was classified as mild, moderate, or severe according to the test instructions. The severity classification on the mColorDx was based on the maximum number of errors made on the red-green classification plates. According to the instructions, less than 18 errors was classified as a mild defect, between 18 and 28 errors was a moderate defect, and 29 or more errors was a severe defect. The Ishihara and PIPIC test instructions stated that the severity of the defect could be qualitatively graded into either mild (no errors on the diagnostic plates) or severe (missing any figure on the diagnostic plates).

RESULTS

Red-Green Screening Series

Fig. 1 shows the kappa coefficient agreement (κ), false negatives, and false positives for all four tests in comparison with the anomaloscope. The κ values for each test were at least 0.95 and, based on the 95% confidence intervals, were not significantly different. A value of 0.95 indicates that the test agreed with the anomaloscope on 95% of the subjects as to who was color-normal and who was color-defective after correcting for chance agreement.[6] The number of false positives for all printed tests was zero, which indicates that the specificity was 1.00. Only one color-normal failed the mColorDx screening series and so specificity was slightly less at 0.99. The false negatives were due to one deuteranomalous subject who passed all of the tests and one additional deuteranomalous subject who passed the

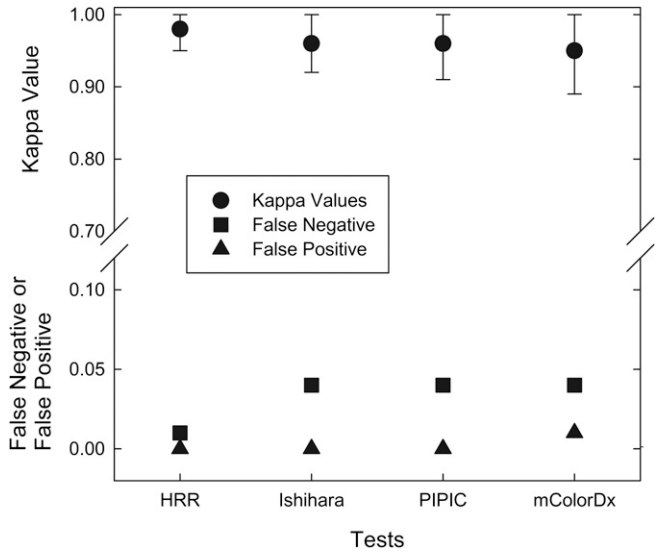

Fig. 1. Kappa coefficients of agreement (circles) with the anomaloscope for the red-green screening series for each test along with false positives (triangles) and false negatives (squares). Error bars are the 95% confidence intervals.

Ishihara, PIPIC, and mColorDx tests. The resulting sensitivities were 0.99 for the HRR and 0.96 for the other tests.

Fig. 2 shows the repeatability of the red-green screening for the tests. All tests showed good agreement between the first and second visit with κ values of 0.96 or greater. The single discrepancy between the two sessions on the HRR was the result of one deuteranomalous subject who failed the screening plates at the first visit but passed at the second visit. On the Ishihara, the single discrepancy was another deuteranomalous subject who failed the first session but passed the second session. This subject also failed the PIPIC and mColorDx tests at the first session, but passed each test at the second session. There was one additional deuteranomalous subject who failed the PIPIC and mColorDx screening plates at the first visit, but passed at the second visit. The color normal who failed the mColorDx at the first visit also failed at the second session. None of the subjects passed at the first session and then failed at the second session.

Blue-Yellow Screening Plates

Table I shows the number of subjects who failed the blue-yellow screening plates on the three tests at each session. The table shows that the deutans were more likely to fail the blue-yellow screening plates, especially on the mColorDx, with 19% of the deutan subjects failing the blue-yellow series at the first session. None of the subjects failed just the PIPIC blue-yellow screening series. As expected, all the individuals who failed the mColorDx blue-yellow plates missed the most desaturated test plates, but some subjects also made errors on the highest saturated plates without making errors on the midsaturated test plates.

Failures on the blue-yellow plates were not always repeatable. The deuteranomalous subject who failed all three blue-yellow screening tests at the first session passed all three at the second session. The other deuteranomalous subject, who failed the HRR and mColorDx at the first session, passed both at the second session. The deuteranomalous who failed both the HRR and mColorDx at the second session failed just the mColorDx at the first session. All of the subjects who failed the mColorDx at the second session failed the test at the first session.

Age-related changes could not entirely explain all the blue-yellow failures because 60% of the subjects who failed the blue-yellow mColorDx series at the first visit were younger than 30 yr. Nevertheless, the older deutan subjects were more likely to repeat a blue-yellow failure. Of the five subjects who failed any blue-yellow screening series at the second session, 67% were older than 30 yrs.

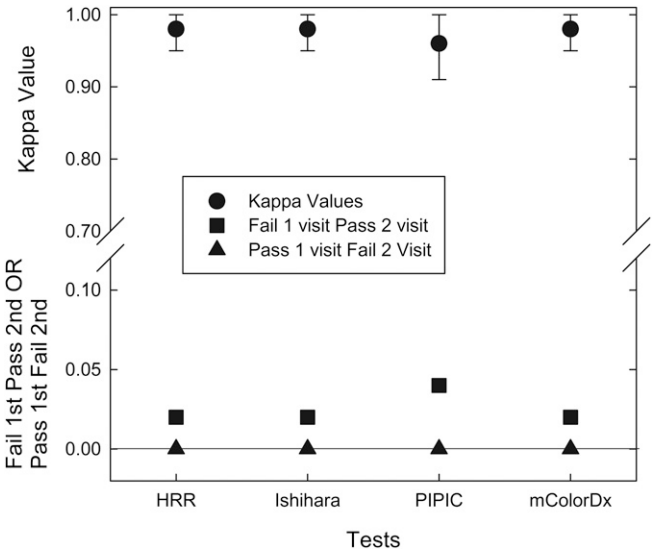

Fig. 2. Kappa agreement values (circles) for passing or failing the screening tests at two different sessions along with the between session discrepancies of failing the 1st session and passing the 2nd session (squares), and passing the 1st session and failing the 2nd session (triangles). Error bars are the 95% confidence intervals.

Diagnostic Plates

Table II shows classification results for all the tests at the first visit as a function of the different types of color vision defects. There were no misclassifications on any of the tests. Both the HRR and mColorDx tests correctly classified 100% of the red-green color-defectives as either protan or deutan at both visits. The color normal who failed the screening on the mColorDx test is not included in the table. He had an equal number of errors on the deutan and protan series at both sessions and so was diagnosed as unclassified. Because this person had normal color vision, but failed the mColorDx screening plates, the unclassified mColorDx diagnosis could be considered the correct diagnosis.

The Ishihara test correctly classified 76% of the deutans and 86% of protans at the first visit. The remaining subjects were unclassified because they did not make any errors on the diagnostic plates. None of the subjects was unclassified because they missed both figures on the diagnostic plates. The classification results for the PIPIC test were similar to the Ishihara test with 77% of the deutans and 86% of the protans classified correctly by the diagnostic plates at the first visit. The color-defective subjects who were unclassified either missed both numbers on the plate (four deuteranopes and one deuteranomalous) or read both numbers on the plates correctly (six deuteranomalous and two protanomalous). Because the HRR agreement with the anomaloscope was perfect in terms of classifying the defect as either deutan or protan the κ coefficient was equal to 1.0 with a standard error of zero.

The κ coefficients for the repeatability classifications for those who failed both sessions were lower and equal to 0.77 (95% CI 0.61 to 0.94) for the PIPIC and 0.84 (95% CI 0.70 to 0.99) for the Ishihara test. Based on the 95% confidence interval, both values were significantly less than 1.00. The discrepancies between both visits on the Ishihara test were due to 2% of the color-defectives being unclassified at the first visit, but classified correctly at the second visit (all protans) and 7% of the color-defectives classified as deutan at the first visit but were unclassified (saw both figures) at the second visit. The discrepancies between the first and second visit for the PIPIC test were due to 12% of the color-defectives who were unclassified at the first visit, but were classified correctly as deutan at the second visit, and 2% of the color-defectives who were classified as

Table I. Number of Individuals Who Failed the Tritan Screening Plates on the Different Tests in the First and Second Sessions.

	ONLY mCOLORDX	mCOLORDX & HRR	ALL 3 TESTS (mCOLORDX, HRR, PIPIC)
1st Session	1 CVN* 2 DA & 5 D	1 DA	1 DA
2nd Session	1 DA & 3 D	1 DA	

* CVN is color-normal; DA is deuteranomalous; D is deuteranope.

deutan at the first visit but were unclassified at the second visit. There were no discrepancies where a person was classified as a protan on one visit and deutan at the second visit or vice versa on either test. Interestingly, 83% of the color-defectives who were unclassified by the Ishihara at the first session remained unclassified at the second session. For the PIPIC, this percentage was lower at 55%.

Severity

Fig. 3 shows the agreement between the Ishihara and PIPIC test in terms of classifying the defect as either mild (i.e., no errors on the diagnostic plates) or severe (any error on the diagnostic plates) at the first session. All the dichromats were classified as a severe red-green defect on both tests. The κ value for agreement on classification for the two tests was 0.66 and was significantly less than 1.0 based on the 95% confidence interval. The discrepancies were four deuteranomalous subjects who were classified differently by each test.

There was a fair agreement for the Ishihara severity classification between sessions for those subjects who failed both sessions. The κ value for repeatability was 0.66 (95% CI 0.37 to 0.95). The discrepancies in the repeatability of the Ishihara were two subjects (protanomalous and deuteranomalous) who were classified as mild in the first session, but were classified as severe in the second session, and two subjects (deuteranomalous and deuteranope) who were classified as severe in the first session, but were classified as mild in the second session. The repeatability for the PIPIC severity grade was good, with a κ equal to 0.81 (95% CI 0.56 to 1.00). One discrepancy was a deuteranomalous subject who was classified as mild in the first session, but severe in the second session. This subject also had the same discrepancy on the Ishihara test. The second discrepancy was a deuteranope who was classified as severe in the first session but mild in the second session. This person was classified as severe by the Ishihara at both visits.

Fig. 4 shows the relationship between the HRR severity classification and the range of acceptable matches for the anomaloscope at the first visit. The mean values from Cole et al.'s study are included for comparison.[3] The figure shows an increase in the anomaloscope range with the HRR severity rating. The Spearman rank correlation coefficient of 0.6 was only moderate, but significant ($P < 0.0001$). Although none of the dichromats was classified as mild, 27% (14% of the deuteranopes and 62% of the protanopes) were classified as moderate instead of severe. There were also two deuteranomalous subjects who would be classified as mild based on a relatively small matching range (i.e., < 20 units) who were classified as moderate or severe on the HRR.

There was a fair agreement in the severity classification of HRR for the first and second visit with a κ coefficient of 0.73 (95% CI 0.55 to 0.91). Seven (15.6%) subjects had a different severity classification at the two visits. Four of these individuals (three dichromats and one anomalous trichromat) improved from severe to moderate upon repetition. The other three subjects (two deuteranomalous and one deuteranope) were classified as more severe in the second session by one level.

Fig. 5 shows the relationship between the mColorDx severity classification and the matching ranges of the anomaloscope at the first visit. The color-normal who failed is excluded from the figure. There was an increase in mColorDx severity grade with increasing anomaloscope matching ranges. The Spearman rank correlation coefficient of 0.45 was significant ($P = 0.002$), but lower than the HRR results. In contrast to the HRR, one deuteranope was classified as mild. He was classified as severe on all the other tests. In addition, there were two deuteranomalous individuals with relatively small anomaloscope match ranges (i.e., < 20) who were classified as severe by the mColorDx test.

The repeatability of the mColorDx test for severity was good, with a κ coefficient of 0.86 (95% CI 0.77 to 0.95). Two color-defectives had discrepancies in their severity. One protanomalous was classified as severe at the first session, but was classified as moderate at the second visit. One deuteranomalous was classified as mild during the first visit, but classified as moderate during the second visit. The color-normal who failed the mColorDx was classified as mild on both visits. His

Table II. Percentage of Subjects Who Were Classified as Deutan or Protan by All Tests in Comparison with the Nagel Anomaloscope Diagnosis.

	HRR		ISHIHARA		PIPIC		mCOLORDX	
ANOMALOSCOPE DIAGNOSIS	CORRECTLY CLASSIFIED	UNCLASSIFIED	CORRECTLY CLASSIFIED	UNCLASSIFIED	CORRECTLY CLASSIFIED	UNCLASSIFIED	CORRECTLY CLASSIFIED	UNCLASSIFIED
D	100%	0	100%	0	78%	22%	100%	0
DA	100%	0	55%	45%	28%	72%	100%	0
P	100%	0	100%	0	100%	0	100%	0
PA	100%	0	67%	33%	67%	33%	100%	0
κ coefficient (95% CI)	1.00 (1)		0.73 (0.55 to 0.91)		0.53 (0.35 to 0.70)		1.00 (1)	

D is deuteranope; DA is deuteranomalous; P is protanope; and PA is protanomalous.

		Ishihara	
		Mild	Severe
PIPIC	Mild	5	2
	Severe	2	36

κ= 0.66 (95% CI 0.36 to 0.96)

Fig. 3. Comparison between the Ishihara and PIPIC test severity classifications for the first session.

mild classification was consistent with his range of acceptable matches of 5 units (39 to 44). The deuteranope who was classified as mild in the first visit was also classified as mild in the second visit.

DISCUSSION

The current study confirms that all four pseudoisochromatic tests are valid red-green color vision screening tests. Although not statistically significant based on the 95% confidence intervals, the HRR had the highest level of agreement with the anomaloscope and the highest sensitivity. The HRR did misclassify one mild deuteranomalous individual as color-normal, but this person was also classified as normal by the other three tests. The Ishihara and PIPIC test screening results were identical, with the same deuteranomalous subjects passing both tests and none of the color-normals failing. The slightly lower agreement between the anomaloscope and the mColorDx was a result of the previously mentioned two deuteranomalous subjects who passed the test in addition to the one normal subject who failed this test. The sensitivity values for the HRR and Ishihara tests were nearly identical to values reported by Cole et al. and Birch in their studies. However, our specificity values for these two tests were slightly higher than found in both studies.[2,3] The slight difference could be because the earlier studies included subjects as young as 8 yr. The younger color-normals may have been more likely to fail these tests.

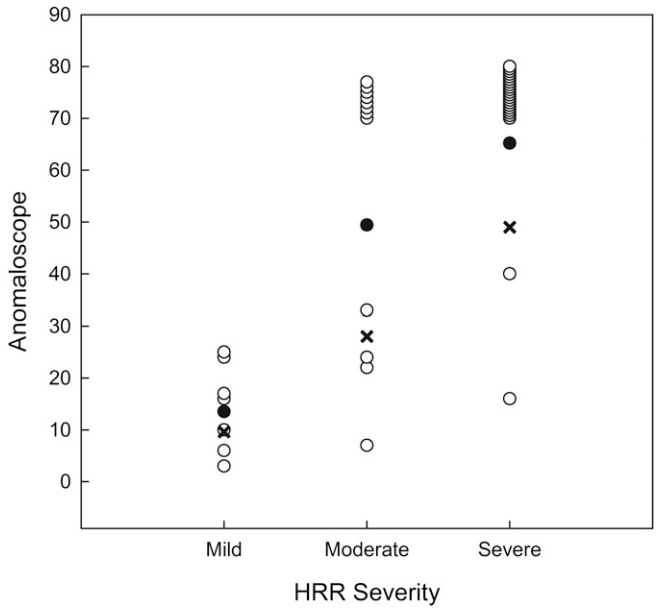

Fig. 4. Relationship between the HRR severity classification and the anomaloscope range of acceptable matches for the color defective subjects (individual dichromats' results have been offset to show the number of individuals at each severity grade). Black circles indicate the average matching range for the anomaloscope in this study. The black Xs are the mean values from Cole et al.'s study.[3]

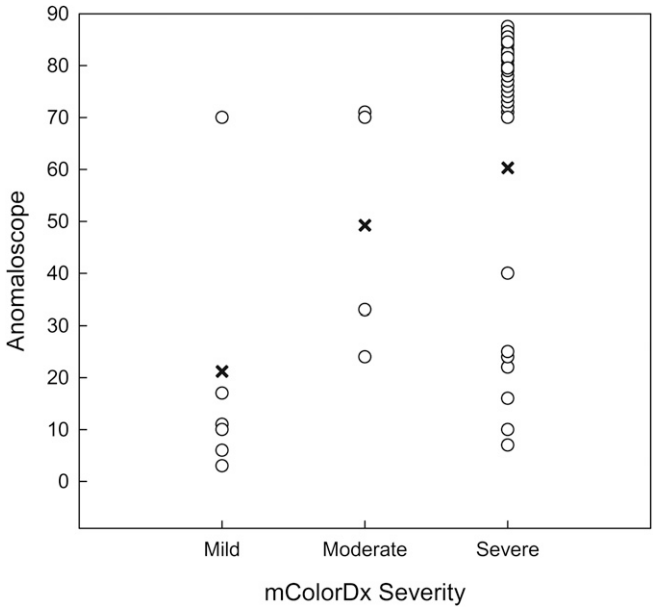

Fig. 5. Relationship between the mColorDx severity classification and the anomaloscope matching ranges for color defective subjects (dichromats' individual ranges have been offset to show the number of individuals at each severity grade). Xs are the average anomaloscope matching ranges for each mColorDx grade.

The κ coefficient for all tests in terms of the screening repeatability was 0.96 or more and statistically identical based on the 95% confidence intervals. All of the dichromats failed the red-green screening plates in all four tests at the both visits. The repeatability of the Ishihara was similar to the values of earlier studies summarized by Working Group 41 even though a different scoring criterion may have been used.[11]

The agreement of the diagnostic plates with the Nagel anomaloscope in terms of the type of red-green defect varied from perfect with the HRR and mColorDx tests to fair for the Ishihara and PIPIC tests. Classification was correct on both the Ishihara and PIPIC tests for approximately 76% of the deutans and 86% of the protans. The lower agreement for these tests was likely due to the limited numbers of the diagnostic plates in both tests combined with our procedure of not asking the subjects which figure was more visible if they reported both figures on the diagnostic plates. Despite the different scoring procedure for the classification plates on the Ishihara, these values are similar to the percentages reported by Birch.[2] She found that 83% of the protans and 82% of the deutans were correctly classified based on either which figure

was missed or which figure was more visible if they reported both figures.

Our HRR classification results were slightly better than reported by Cole et al.[3] There was 100% agreement in our study as to whether the color-defective subjects were classified correctly as protan or deutan, whereas Cole at al. reported that only 86% of the color-defectives were correctly classified as protan or deutan. This may have been a result of their color-defective sample having a higher percentage of individuals with milder defects. These individuals may have been less likely to miss any diagnostic figures or they missed an equal number of protan and deutan plates.

In terms of the repeatability of the classification, the agreement between the first and second visit for classification was perfect for the HRR and mColorDx tests, good for the Ishihara test, and fair for the PIPIC test. Again, the limited number of plates in the PIPIC and Ishihara was likely responsible for the lower repeatability.

All four tests can qualitatively classify the severity of the defect. The Ishihara and PIPIC tests grade the severity as mild or severe, whereas the HRR and mColorDx tests were designed to classify the severity as mild, moderate, or severe. Our results from the HRR test were similar to Cole et al.'s in that there was a reasonable correlation between the HRR severity and the Nagel anomaloscope matching ranging. Similar to previous studies, none of the dichromats were classified as mild.[1,3] Our results were also similar to Cole et al.'s finding that a small percentage of dichromats are classified as moderate and a small percentage of individuals with a mild defect based on the anomaloscope range were classified as moderate.[3] The reason that our mean matching ranges in Fig. 4 were higher than Cole et al.'s values was probably because they had a higher percentage of color-defective subjects with milder defects.

The mColorDx test ability to grade the severity was slightly lower than the HRR plates. This was due to one dichromat being classified as mild, two dichromats being classified as moderate, and two deuteranomalous with a relatively small matching range being classified as severe in the first session. Interestingly, these two individuals with the smaller matching range were also classified as moderate by the HRR test. However, there were also two deuteranomalous subjects who were classified as mild by the mColorDx but as moderate by the HRR test.

In terms of the repeatability of the severity classification, none of the subjects moved more than one severity category between sessions. The κ coefficient for repeatability of the mColorDx was better than the HRR. However, the mColorDx is more likely to classify a dichromat as mild and this mistake is repeatable.

Although the number of subjects was limited, our results suggest that a person who fails the mColorDx red-green screening plates and has an unclassified mild defect could be a false positive. Ideally, one would use the anomaloscope to make a definitive diagnosis, but either the HRR or Ishihara would be sufficient.

The HRR, PIPIC, and mColorDx tests can screen for blue-yellow defects and a number deutan subjects failed these screening plates, especially on the mColorDx. The small number of failures on the printed tests is unlikely to be repeated, whereas about 50% of the subjects are likely to repeat a failure on the mColorDx test. Bailey et al. also reported that red-green color defectives occasionally have an error on the blue-yellow plates on the HRR.[1] We believe that there are multiple reasons for the blue-yellow failures on the mColorDx. The first is that the deutan subjects may have a discrimination ellipse/zone of confusion around the gray background that is either wider than assumed when designing the blue-yellow test plates or the zones/ellipses of these individuals are rotated slightly toward the blue-yellow lines of confusion. The widening or rotation could be a result of age-related changes, differences in ocular media in younger subjects, differences in the photoreceptor pigments, or any combination of these factors. We do not believe that age-related changes are solely responsible since half of the subjects were under 31 yr at the first session. Nevertheless, individuals who are older or have a deuteranopic defect are more likely to repeat the failure on the blue-yellow plates.

The second reason for an increased number of errors on the blue-yellow plates could be the short presentation time. The 2-s presentation time used by the mColorDx may be too brief for individuals who also have a reduction in sensitivity in the red-green dimension. A 2-s presentation time has been shown to result in more errors on printed blue-yellow screening plates than red-green screening plates for color-normals and so it is possible that the effect of shorter presentation times is greater for individuals with a red-green defect even though the test colors are approximately orthogonal to their red-green axis of confusion.[14]

Third, it is also possible that these mistakes could be an error of expectation in that the subject was expecting not to perceive any figure or symbol following the red-green screening plates and did not examine the subsequent blue-yellow test plate carefully. This type of error was noted on the Standard Pseudoisochromatic Plates Part 2 with younger and older color-normals more prone to making this error.[7] Although none of the color-normals missed the first two saturated blue-yellow test plates, 15% of the color-defects missed at least one of these plates in the first session and then entered correct responses for the midsaturation plates. This type of response pattern supports the hypothesis that some of the subjects were not expecting to see a figure when the blue-yellow screening started. The mColorDx only provides the responses on the individual plates as correct or incorrect and so it was not possible to determine the nature of their incorrect responses. Regardless of the underlying case, the resulting mixed defect makes interpreting the test results more difficult in the screening environment where other clinical findings may be unavailable.

One of the potential drawbacks of repeating the tests in the same order is that any fatigue effects associated with each test, particularly the mColorDx, which was the last test in the sequence, would be carried over into the second session instead

of being randomized. However, our results showing that 50% of the subjects who failed the mColorDx blue-yellow plates in the first session, but passed in the second session, suggest that any fatigue effects were masked by practice effects. This finding is consistent with a previous report that performance on clinical color vision tests is robust to fatigue resulting from sleep deprivation.[8]

The four pseudoisochromatic plate tests (HRR, Ishihara, PIPIC, and mColorDx) can be used confidently to detect red-green color vision defects. Although the level of agreement with the anomaloscope was statistically identical, the HRR may be preferred over the Ishihara, PIPIC, and mColorDx because its sensitivity was marginally higher than the other three tests without any tradeoff in the specificity. All four screening tests were highly repeatable screening tests for red-green defects.

The HRR, PIPIC, and mColorDx tests can also screen for blue-yellow defects. Deutan individuals were prone to make errors on these plates, especially on the mColorDx. The blue-yellow failures were likely caused by multiple factors, including age, presentation times, and errors of expectation and entering incorrect responses.

The HRR and mColorDx were the most effective in classifying the defect as either protan or deutan. One possible reason for the better performance of the HRR and mColorDx was that they presented more diagnostic plates compared with either the Ishihara or PIPIC tests.

Both the HRR and mColorDx tests provided a better qualitative diagnosis of the severity of the defect compared with the PIPIC and Ishihara. However, the agreement with the anomaloscope was not perfect, with some dichromats classified as mild or moderate and individuals with a mild defect classified as moderate or severe. The mColorDx was more prone to these misclassifications. This result suggests that the HRR and mColorDx should not be used as a single test in determining the severity of the defect. Other tests, such as the Farnsworth D15 test or the anomaloscope, should be administered if determining the severity of the defect is important for occupational reasons.

ACKNOWLEDGMENTS

The Military ColorDx test was graciously supplied by Dr. Terry Waggoner.

King Saud University and the Saudi Arabia Cultural Bureau in Ottawa provided financial support to Ali Almustanyir.

Authors and affiliations: Ali Almustanyir, B.Sc. Optom., M.Sc., and Jeffrey K. Hovis, O.D., Ph.D., School of Optometry and Vision Science, University of Waterloo, Waterloo, Ontario, Canada; and Ali Almustanyir, B.Sc. Optom., M.Sc., College of Applied Medical Sciences, Optometry Department, King Saud University, Riyadh, Saudi Arabia.

REFERENCES

1. Bailey JE, Neitz M, Tait DM, Neitz J. Evaluation of an updated HRR color vision test. Vis Neurosci. 2004; 21(3):431–436.
2. Birch J. Efficiency of the Ishihara test for identifying red- green colour deficiency. Ophthalmic Physiol Opt. 1997; 17(5):403–408.
3. Cole BL, Lian KY, Lakkis C. The new Richmond HRR pseudoisochromatic test for colour vision is better than the Ishihara test. Clin Exp Optom. 2006; 89(2):73–80.
4. Dain SJ. Clinical colour vision tests. Clin Exp Optom. 2004; 87(4-5): 276–293.
5. Deeb SS. Molecular genetics of colour vision deficiencies. Clin Exp Optom. 2004; 87(4-5):224–229.
6. Hays WL. Statistics. New York: Holt, Rinehart and Winston. Inc.; 1988.
7. Hovis JK, Cawker CL, Cranton D. Normative data for the standard pseudoisochromatic plates–Part 2. J Am Optom Assoc. 1990; 61(12):913–920.
8. Hovis JK, Ramaswamy S. Color vision and fatigue: an incidental finding. Aviat Space Environ Med. 2007; 78(11):1068–1071.
9. Lakowski R. Theory and practice of colour vision testing: a review 1. Br J Ind Med. 1969; 26(3):173–189.
10. Pokorny J, Smith VC, Verriest G, Pinckers AJLG, editors. Congenital and acquired color vision defects. New York: Grune & Stratton; 1979.
11. Report of Working Group 41. Procedures for testing color vision. Committee on Vision, Assembly of Behavioral and Social Sciences, National Research Council. Washington (DC): National Academy Press; 1981.
12. Schmidt I. Some problems related to testing color vision with the Nagel anomaloscope. J Opt Soc Am. 1955; 45(7):514–521.
13. Schneck ME, Haegerstrom-Portnoy G, Lott LA, Brabyn JA. Comparison of panel D-15 tests in a large older population. Optom Vis Sci. 2014; 91(3):284–290.
14. Somerfield MR, Long GM, Tuck JP, Gillard ET. Effects of viewing conditions on standard measures of acquired and congenital color defects. Optom Vis Sci. 1989; 66(1):29–33.