MCSL

Munsell Color Science Laboratory

Technical Report

Evaluation and Correlation of Color Discrimination Abilities

Sun Ju Park
Scot R. Fernandez
Lawrence Taplin
Munsell Color Science Laboroatory
Chester F. Carlson Center for Imaging Science
Rochester Institute of Technology
54 Lomb Memorial Drive
Rochester, New York 14623-5604 USA
(716)475-2784
Fax: (716)475-5988

June, 2000

This study had two parts first evaluating the variability of color discrimination determined by the Colorcurve HVC Color Vision Skill test within a population of "normal" observers and a colorimetric evaluation of the ColorCurve Scoring. For the first evaluation, *more than normal variance was found on the red and yellow rows of color. No statistical mean difference was found between any of the observer characteristics noted, such as age, gender, experience, or vision ability (corrected/uncorrected). Finally, this evaluation compared the results of two different color discrimination tests; the first population used the Colorcurve HVC Color Vision Skill test and the second population used the Farnsworth-Munsell 100 Hue test. The results found no statistical correlations between the two test methods. The second evaluation is discussed in the last part of the appendix, but the understanding is that scoring does not necessarily equate colorimetrical.*

I. INTRODUCTION

It is understood that human color vision demonstrates variability both within an individual observer (inter-observer) and from observer to observer (intra-observer). These variabilities demonstrated by the human visual system can be broken down into two primary categories, color vision deficiency and normal color vision. This study is not interested in color vision defectiveness, but in better evaluating the variability of color discrimination within a population of "normal" observers. Color discrimination is the measurement of how well an observer can differentiate between colors. Several tests have been developed to quantify this ability of color discrimination. The Colorcurve HVC Color Vision Skill test is evaluated for its ability of quantifying its color discrimination, and the Farnsworth Munsell 100 Hue test is selected to find out whether there is consistent evaluation of the two tests' color discrimination.

The focus of the proposed report is to observe the collected data and to evaluate the results of the aforementioned color discrimination tests done by the Colorcurve HVC Color Vision Skill test. First of all, the sample data of the Colorcurve HVC Color Vision Skill test will be compared to the given evaluation guidelines of the test. Second, means of different groups will be compared to each other in order to find out the correlation between color discrimination and observer characteristics. Finally, the consistent evaluation of the two selected tests will be examined from the correlations between the results. A literature search revealed that a previous comparison of testing methods found no correlation between tests. Re-affirmation of the findings will be examined as the last objective of the report, which evaluates the existence of any relationship between the two tests. If there is a strong relationship between the two tests, the observers' consistent color discrimination can be observed.

II. EXPERIMENTAL

Defining the viewing conditions of any visual experiment is an integral element for reducing random error. The viewing conditions for the Colorcurve HVC Color Vision Skill test are as follows: source equal to illuminant D65, approximating a 0/45 geometrical relationship between the light source, sample, and observer, and observer positioned approximately 16-18 inches from the samples. Each test used the testing procedures as defined by their corresponding manufacturers. 28 observers participated in the Colorcurve HVC Color Vision Skill test, and 13 observers participated in the Farnsworth-Munsell 100-Hue test.

The Colorcurve HVC Color Vision Skill test consists of four rows of colors: red, yellow, blue, and green. Each row has nine samples, and in a separate carrying case are 36 colored tiles placed face down in a random order. Each tile is taken individually and matched to its best color match according to the observer, and then returned to the carrying case face down. It is possible that the observer could choose more than one match for each color sample, and none for others depending on the observers' decisions. Each tile and sample combination has an optimal pick; however, partial credit is awarded on close matches. After the test is complete, each row is scored and summed up to provide an overall score of color discrimination.

The Farnsworth-Munsell 100-Hue test evaluates color discrimination ability and also provides a quick assessment of some color vision defectiveness such as dichromacy. The test consists of 85 caps, each cap containing a unique color patch. The caps, when organized in the correct order, create a continuous hue circle from red, to orange, yellow, green, blue-green, blue, purple, and back to red. The reverse side of the each patch contains a number placing the patches in the correct order. The observer is presented the test with the color patches in a random order, and does his or her best to reconstruct the correct order based on continuous color. Once the re-organization of color patches is complete, a comparison of the observed order versus the theoretical order is made and the discrimination level of the subject can be assessed.

III. RESULTS AND DISCUSSION

The general descriptive statistics are computed in order to understand the general results of the vision tests (Table I). The Colorcurve HVC Color Vision Skill test tells how good color discrimination is for a specific color. On the other hand, the Farnsworth-Munsell 100-Hue test tells how bad color discrimination is for a specific color. In order to apply statistical and mathematical computations to the sample, which often requires the normal distribution of the data, a normality check of the samples collected by the two tests are done by using the Q-Q plots shown in Figure 1 in Appendix. Once normality is determined, further statistical techniques and interpretations are made based on the normal distribution.

The Colorcurve HVC Color Vision Skill test

Variable	N	Mean	StDev	Minimum	Maximum
Total Score	28	60.50	13.61	26	78

Blue	28	17.14	5.69	3	25
Red	28	13.93	4.13	5	22
Green	28	15.14	4.47	2	22
Yellow	28	14.29	3.74	5	21

The Farnsworth-Munsell 100-Hue test

Variable	N	Mean	StDev	Minimum	Maximum
Total Score	13	13.69	17.12	4	66
Red	13	1.92	5.79	0	21
Green	13	4.92	8.08	0	29
Blue	13	2.38	3.84	0	11
Purple	13	4.46	3.10	0	11

Table I. General Descriptive Statistics of Two Tests

Three questions are set at the beginning to direct the evaluation of the Colorcurve HVC Color Vision Skill test.

- Does the sample data of the Colorcurve HVC Color Vision Skill test fall within the given range of the manufacturer's guidelines?
- Do the different populations report different color discrimination abilities?
- Are the Colorcurve HVC Color Vision Skill test and the Farnsworth-Munsell 100 Hue test consistent with each other in evaluating color discrimination?

First Question:

In order to test the multivariate population means of blue, red, green, and yellow, the following procedure is processed based on the hypothesis testing, univariate and multivariate mean vector testing with T^2.

- H_0: All test score means of four colors fall into the 90% confidence intervals of the population test score means of four colors. (μ_i = x_bar$_{i}$, $_{i}$=blue, red, green, yellow)
- H_1: Not all test score means of four colors fall into the 90% confidence intervals of the population test score means of four colors. ($\mu_i \neq$ x_bar$_{i}$, $_{i}$=blue, red, green, yellow)

The maximum score of the color vision test is 100. Each score of the color contributes approximately 25% of the total score. Based on the test guidelines, the average score range of color discrimination for normal vision is from 58 to 69. The average of the two extreme values in the average range is 63.5. Based on the calculation, 63.5 : 100 = *x* : 25, the theoretical average of the average range for each color is about 16 (rounded up from 15.88). The hypothesis testing is done by setting the theoretical average to 16 and at the 10% level of significance. According to the T^2 test (Equation1), H_0 is rejected.

Equation1. If $T^2 = n(\bar{x} - \mu)' S^{-1} (\bar{x} - \mu) > (n-1)p/(n-p) F_{p,n-p}(\alpha)$, then reject Ho.

(T^2) 16.42 > 9.88 (F Distribution)

This rejection of H_0 indicates that one or more of the multivariate means, or some combination of means differ too much from the hypothesized mean values of [16,16,16,16]. In order to find out the cause of the rejection, univariate calculations for each color are done using the t^2 test. The results are as follows in Table II.

	t^2		t distribution (0.1 level of significance)	
Blue	1.13	<	2.90	Cannot Reject Ho
Red	7.05	>	2.90	Reject Ho
Green	1.03	<	2.90	Cannot Reject Ho
Yellow	5.88	>	2.90	Reject Ho

Table II. Univariate Analysis for Each Color

It turns out that the test scores of red and yellow colors are too far away from their hypothesized averages. This indicates that large variances of the colors red and yellow are noticed (more than normal variance given by the test guidelines). It is concluded that the range for judging the average color discrimination provided by the test guidelines is high compared to the sample data. The possible solution for adjusting the scoring of the Colorcurve HVC Color Vision Skill test is either lowering the average of the average range of normal color discrimination or less weighting two colors, red and yellow, which are evaluated as being too far from the theoretical mean.

Second Question:

Mean vectors from two populations based on observer characteristics, which include age, gender, experience level, and vision ability, are compared. These characteristics are tracked so that any correlation between observer traits and color discrimination could be observed. In order to compare multivariate means of different groups, the following procedure is processed based on the hypothesis testing and balanced MONOVA (Multivariate Analysis of Variance).

- H_0: Mean scores of four colors from two groups are all the same.
- H_1: Mean scores of four colors from two groups are not all the same.

Based on the results of balanced MONOVA, there is no mean difference between two different populations such as experienced vs. non-experienced groups, 20's vs. 30's age groups, female vs. male groups, and eye correction vs. no eye correction groups (Table III). In other words, the variability of color discrimination is not determined by specific characteristics of observers. This agrees with the test guidelines, which does not mention about specific traits of the observer when evaluating his or her color discrimination.

Response \ Factor	Experience vs. No Experience	Age Difference	Gender Difference	Vision Ability
Blue	0.61	0.26	0.41	0.61
Red	0.9	0.65	0.23	0.19
Green	0.32	0.21	0.34	0.09
Yellow	0.71	0.51	0.65	0.09
CRITERION	Experience vs. No Experience	Age Difference	Gender Difference	Vision Ability
Wilk's	0.91	0.44	0.77	0.12
Lawley-Hotelling	0.91	0.48	0.77	0.13
Pillai's	0.91	0.41	0.77	0.11
Roy's				

Table III. Balanced MONOVA

Third Question:

Finally, of interest to observe is how well both tests correlated to each other when evaluating the color discrimination. The correlation matrix between the same colors evaluated by the two tests is given in Table IV. The strongest correlation exist between greens, which scores –0.88. This means that if the observer earns high color discrimination ability on the green color by the Colorcurve HVC Color Vision Skill test, but low color discrimination ability on the green color by the Farnsworth-Munsell 100 Hue test. The negative correlation between the two tests indicates that the different scaling of the measurements, one is how good, and the other is how bad on color discrimination. Since not obvious correlations between other colors are found between two test colors, the agreement with the prior-report is made.

	Blue(Colorcurve)	Red(Colorcurve)	Green(Colorcurve)	Yellow(Colorcurve)
Blue(100 Hue)	-0.35			
Red(100 Hue)		-0.46		
Green(100 Hue)			-0.88	
Purple(100 Hue)				0.28

Table IV. Correlation between Two Vision Tests

IV. CONCLUSIONS

There are two improvements needed in order to support the evaluation of the Colorcurve HVC Color Vision Skill test. First of all, when collecting the data from the HVC Colorcurve test, the test setting was not the best to generate the proper data set. Even though the experiments followed the given viewing and illumination conditions, stray lights and semi-reflection disturbed the observers when they did perform the color match. More careful setting by using a black curtain, which may eliminate stray lights of the surrounding, is desirable. This may higher the scores of the sample close to the given average range of the test. To reduce the uncertainty of the results, several retests are necessary to support the conclusion. Secondly, 13 sample data collected by the two tests were used when producing the correlation matrix between the tests, which data set may be too small to generalize the results. More experiment data is needed to confirm the findings.

V. REFERENCES

Richard A. Johnson and Dean W. Wichern, Applied Multivariate Statistical Analysis, Fourth Edition, Prentice-Hall, Inc., 1998.

Colorcurve HVC Color Vision Skill Test User's Manual.

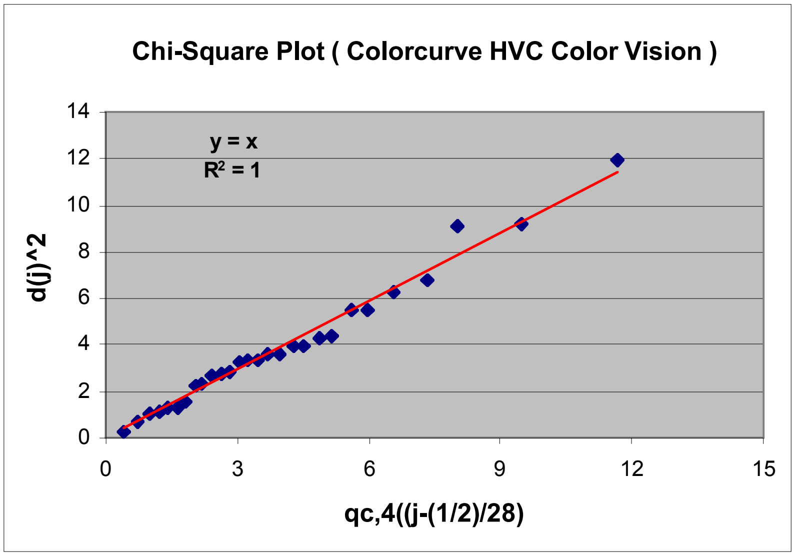
The Farnsworth-Munsell 100-Hue Test for the examination of Color Discrimination.

IV. APPENDICES

Uncorrected = 0
Glasses = 1
Contacts = 2

					HVC Color Curve					Farnsworth-Munsell				
										Red	Green	Blue	Purple	
Observer #	**Age**	**Gender**	**Eyes**	**Experience**	**Blue**	**Red**	**Green**	**Yellow**	**Total**	**85 - 21**	**22 - 42**	**43 - 63**	**64 - 84**	**Total**
1	27	M	1	1	14	15	16	13	58	0	4	0	0	4
2	19	M	1*	0	7	13	12	12	44					
3	22	M	0	0	15	11	16	20	62	0	4	11	5	20
4	24	M	0	1	24	22	15	15	76	0	0	0	4	4
5	23	F	2	1	16	16	21	19	72					
6	26	F	0	0	24	21	18	10	73	0	0	0	4	4
7	22	M	1	1	3	5	13	5	26					
8	37	M	1	1	18	15	19	17	69					
9	27	F	2	1	21	11	18	15	65	0	4	0	4	8
10	26	F	1	0	9	12	12	12	45	3	12	1	8	24
11	23	M	0	1	22	13	18	10	63	0	0	4	4	8
12	22	M	0	0	8	8	2	11	29	21	29	8	8	66
13	29	M	0	1	23	20	16	16	75	0	0	0	4	4
14	44	F	1	1	14	19	11	17	61					
15	27	M	1	1	14	14	16	15	59					
16	36	M	2	1	15	19	22	21	77					
17	35	M	1	1	16	11	15	10	52					
18	27	F	1	1	25	14	20	18	77					
19	28	M	0	0	21	13	18	14	66	0	4	0	0	4
20	21	M	1	0	16	14	13	14	57	0	0	0	4	4
21	38	M	0	1	15	18	15	19	67	1	0	0	11	12
22	32	M	1	1	16	8	16	13	53					
23	25	M	1	1	18	9	7	12	46					
24	23	M	1	1	20	11	11	10	52	0	7	7	2	16
25	30	M	1	0	25	16	18	19	78					
26	32	F	1	0	22	16	16	13	67					
27	38	M	1	1	23	11	21	13	68					
28	24	M	0	1	16	15	9	17	57					

Table V. Color Discrimination Raw Data



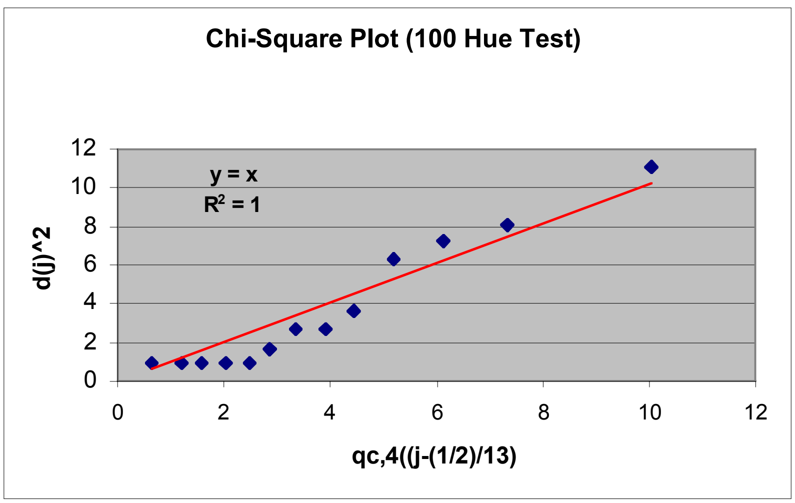


Figure 1. Normality Check of the Samples

Appendix II - Correlation of Instrumental Measurements and ColorCurve Scoring

Measurement Procedure:

Each Colored chip and reference patch of the ColorCurve HVC Color Vision Skill Test was measure five times using a Gretag SpectroLino spectrophotometer. The instrument converted the captured reflectance data from its 45/0 geometry to Lab and XYZ values using the 1931 2-degree color matching functions and a D65 illuminant. The chromaticity coordinates were averaged and are plotted below as a general reference to the area of color space used in the test:

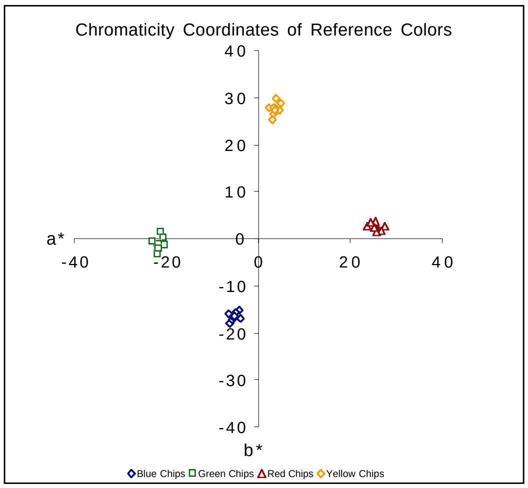


Figure 1 - Chromaticity Coordinates of ColorCurve Reference Colors

Results and Discussion:

The ColorCurve Scoring Data was entered into a spreadsheet and the instrumental color difference for each possible placement of the colored chips into their corresponding row were calculated using ΔE^*_{ab} and ΔE^*_{94}. When evaluating the resulting table below the following points should be kept in mind.

- The best match for a chip does not always receive the maximum 3 points.
- In each color row there is a chip that always receives at least one point for every position.
- Some chips have no positions worth 2 points.
- According to the score sheet, for placing Red Chip #33 in position 6 (from the left) of the Green row the subject receives 1 point. This is an obvious mistake. While evaluating the test, the 33 corrected by changing it to a 36.

These points help to explain why the table values aren't always ordered within the rows by ascending score.

Table 1 - Instrumental Color Differences for ColorCurve Scoring Data

		ΔE*ab					ΔE*94				
		Points					Points				
		0	1	2	3	All	0	1	2	3	All
Blue	**Min**	0.80	0.15	0.74	0.09	0.09	0.76	0.13	0.49	0.07	0.07
	Max	3.59	1.99	1.13	0.18	3.59	2.35	1.44	0.81	0.16	2.35
	Mean	2.01	1.40	0.99	0.14	1.63	1.48	1.05	0.65	0.12	1.20
	StdDev	0.55	0.50	0.12	0.03	0.77	0.36	0.35	0.09	0.03	0.55
Green	**Min**	1.27	0.10	1.13	0.05	0.05	0.76	0.10	0.85	0.03	0.03
	Max	4.88	2.48	3.90	0.16	4.88	3.67	1.86	2.81	0.16	3.67
	Mean	2.31	1.63	2.82	0.08	2.03	1.75	1.20	1.94	0.07	1.52
	StdDev	0.75	0.69	1.13	0.04	1.03	0.60	0.55	0.88	0.04	0.79
Red	**Min**	0.97	0.19	2.26	0.12	0.12	0.47	0.17	1.48	0.09	0.09
	Max	5.28	2.89	3.02	0.32	5.28	5.27	2.89	1.71	0.27	5.27
	Mean	2.61	1.66	2.61	0.20	2.19	2.05	1.26	1.59	0.16	1.67
	StdDev	0.90	0.69	0.34	0.07	1.08	1.05	0.81	0.09	0.06	1.07
Yellow	**Min**	1.30	0.13	1.29	0.06	0.06	0.93	0.11	0.66	0.03	0.03
	Max	4.69	2.69	1.67	0.37	4.69	2.72	1.44	0.92	0.34	2.72
	Mean	2.47	1.60	1.46	0.16	1.99	1.57	1.02	0.76	0.13	1.25
	StdDev	0.75	0.66	0.13	0.10	0.98	0.35	0.32	0.10	0.09	0.56
All	**Min**	0.80	0.10	0.74	0.05	0.05	0.47	0.10	0.49	0.03	0.03
	Max	5.28	2.89	3.90	0.37	5.28	5.27	2.89	2.81	0.34	5.27
	Mean	2.34	1.58	1.95	0.15	1.96	1.71	1.14	1.23	0.12	1.41
	StdDev	0.77	0.63	0.97	0.08	0.99	0.68	0.56	0.71	0.07	0.79

Correlation of Average Subject Response Color Differences and Color Curve Scoring:
Twenty-nine subject score sheets were entered into a spreadsheet and matched with the corresponding instrumental measurements of the chip and patch at the selected position. In this way the ΔE^*_{94} of each subject match could be calculated. The following correlation matrix is the result of these calculations. The correlation coefficients are negative because as the color curve score increases the average ΔE^*_{94} value decrease. A strong correlation was found between the average ΔE^*_{94} of the matches and the color curve scoring for the Blue, Green and Yellow rows. However, the Red row showed a poor correlation, the reason for this could not be directly determined.

Avg ΔE94 \ Avg Color Curve Score	B	R	G	Y	All
B	**-0.98**	-0.45	-0.48	-0.33	-0.80
R	-0.75	**-0.53**	-0.42	-0.52	-0.76
G	-0.49	-0.37	**-0.96**	-0.29	-0.71
Y	-0.34	-0.44	-0.29	**-0.95**	-0.63
All	-0.85	-0.57	-0.70	-0.63	**-0.93**

Table 2 - Correlation Matrix for ΔE94 and Subject ColorCurve Score Data

The use of a single rows score or average ΔE^*_{94} value as a predictor of the subjects overall score was also examined by creating correlation matrices. The yellow row proved to be the poorest predictor in both cases. Interestingly, it was generally the yellow row that subjects complained about having the most difficulty with.

Avg Color Curve Score

	B	R	G	Y	All
B	1.00	0.48	0.50	0.34	**0.82**
R	0.48	1.00	0.36	0.50	**0.76**
G	0.50	0.36	1.00	0.34	**0.73**
Y	0.34	0.50	0.34	1.00	**0.68**
All	0.82	0.76	0.73	0.68	1.00

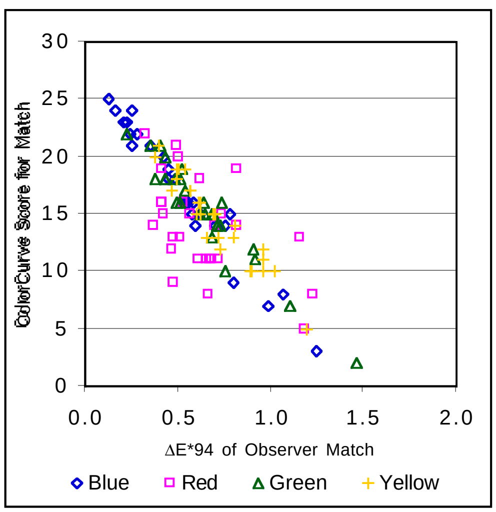
Table 3 - Correlation Matrix showing partial ColorCurve Scores as an indicator of Overall Performance.

Avg ΔE94

	B	R	G	Y	All
B	1.00	0.76	0.46	0.35	**0.85**
R	0.76	1.00	0.42	0.59	**0.89**
G	0.46	0.42	1.00	0.28	**0.70**
Y	0.35	0.59	0.28	1.00	**0.67**
All	0.85	0.89	0.70	0.67	1.00

Table 4 - Correlation Matrix showing average category color differences as an indicator of overall performance.

The data driving the correlation matrix between ΔE^*_{94} and the ColorCurve Scores of 29 observers was also plotted directly. The looser distribution of the red row (squares) is clearly visible.



By sorting the ΔE^*_{94} Color Differences of possible matches first by chip number, then by ΔE^*_{94}, and finally by ColorCurve score in descending, order; 26 reversals between ΔE^*_{94} ordering and the ColorCurve ordering of color matches became evident. These reversals along with the difference in ΔE^*_{94} values are show in the difference in ΔE^*_{94} values are show in the table below.

Color	**Blue**	**Green**	**Red**	**Yellow**	**All**
Count of Reversals	3	10	11	2	26
Min ΔE^*_{94} Difference	0.01	0.05	0.00	0.02	0.00
Max ΔE^*_{94} Difference	0.08	0.91	0.64	0.03	0.91
Average ΔE^*_{94} Difference	0.05	0.34	0.31	0.02	0.27

Table 5 - Reversals of best match ordering between ΔE94 and ColorCurve scoring.

Conceivably, the yellow and blue row reversals were within the precision range of the instrument with which they were made. However, several of the red and green reversals could be confirmed visually by at least one observer. Whether the correction of these reversals by modifying the score sheet would improve the correlation of the red row scores to the average ΔE^*_{94} for actual observer matches was not tested but might prove insightful.