A Comparison of Visual and Spectrophotometric Evaluations of Paired Color Prints

NPIRI Task Force on Color Measurement*

Abstract

Key Words: Color, Print, Spectrophotometer, Visual

Spectrophotometers and derived colorimetric values are being used increasingly to specify and approve color. Yet, in the final analysis, the human eye remains the ultimate arbiter. In order to understand better what the eye sees and what colorimetric values tell us, twenty-six pairs of printed color samples were rated visually by fifteen experienced colorists (8 male, 7 female), and compared with color difference values measured spectrophotometrically.

Visual ratings were made under D50 and Illuminant A lighting using a Pantone® viewing box. Color measurements for each paired sample (batch versus standard) were made with both sphere and 0/45 geometries. A comparison of the visual ratings with DE*, DL*, DC* and Dh values showed in general, good correlation. The degree of correlation varied with hue and the particular color attribute.

Individual panelists were consistent in their ratings. However, significant differences were observed between some of the panelists. Differences were also observed between male and female ratings.

*R. Bassemir
G. Costello
A. DiBernardo
J. DiPiazza
D. Kuna
K. Paulius
C. Rybny
W. Zawacki

Introduction

The adage "beauty is in the eye of the beholder" can be paraphrased to read "color is in the eye of the beholder". Colorimetric measurements notwithstanding, if an art director, advertising manager or anyone else in the graphic arts chain does not agree with the spectrophotometer, the color will not be accepted.

Since the eye is the ultimate arbiter, and since colorimetric measurements are being used increasingly to specify color, it behooves us to learn more about the relationship between what the eye sees, and what colorimetric measurements tell us. The study reported here is one small step toward learning more about that relationship.

Experimental Procedure

Twenty-six pairs of printed color samples which included six hues plus brown and gray were selected for this study. The samples were printed with a laboratory proofer using water base flexographic ink on bleached liner board. The standard and batch were made with the same pigments to minimize metamerism. The batch and standard inks were placed on the anilox roll at the same time so that the two inks were printed side by side. The sample was mounted on a 8" x 8" gray matte paper sheet that had a 2" x 2" hole cut in the center for the sample. The sample number and hue were printed on transparent tape placed at the top of the sheet. The shade of the gray sheet was selected to blend with the shade of the interior of the Pantone® viewing box.

As a precautionary measure, all panelists passed the Ishihara color blindness test. Each panelist viewed the samples in a random order that was selected by computer, first under D50 light and again in the same order under Illuminant A light. Sample identification and viewing order are shown in Appendix Table 1. Panelists sat comfortably in front of the viewing box and were instructed to hold the sample in the center of the box. Panelists were free to hold the sample at any angle in making their evaluation. A standard written form (Appendix Table 2) was completed for each evaluation. All viewings were made with the same viewing box at the same location. Light intensity over the entire visible light spectrum was measured with a spectral radiometer at the start and after all panelists had completed their evaluations (a period of one month) to ensure that there were no significant changes in the light sources.

Questions asked of panelists were as follows:

Is there a color difference?	Yes ___	No ___
Is the batch commercially acceptable?	Yes ___	No ___
Is the batch lighter, darker? ____	No difference ____	
Is the chroma higher, lower? _____	No difference ____	

For the applicable hue:

Is the batch redder, greener, yellower, bluer? ____ ____

In order to quantify visible observations, the following numerical scale was used:

Color Difference
Yes +1
No -1

Batch Acceptance
Yes +1
No -1

Lightness
Lighter +1
Darker -1
No Difference 0

Chroma
Higher +1
Lower -1
No Difference 0

For Orange, Blue, Violet
Redder +1
Greener -1
No Difference 0

For Red
Yellower +1
Bluer -1
No Difference 0

For Green
Bluer +1
Yellower -1
No Difference 0

For Yellow
Greener +1
Redder -1
No Difference 0

For Brown, Gray
Yellower, Bluer, +1
Redder, Greener, -1
No Difference 0

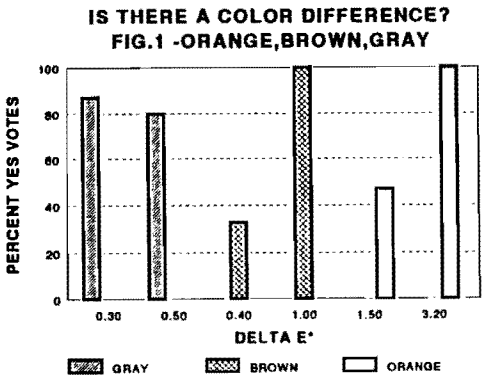
All samples were measured with Datacolor CS-5 and X-Rite 938 0/45 spectrophotometers. Results were reported as CIELAB D50/2 degree observer, and Illum. A/2 degree observer. Complete spectrophotometric measurements are shown in Appendix Table 3.

Experimental Results

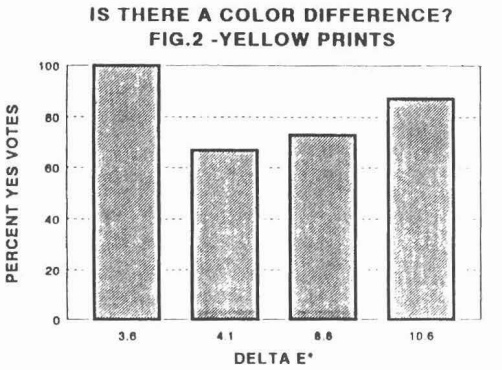
Examination of the colorimetric measurements showed that for these non-glossy samples (85° Gloss less than 4.8) there was very little difference in L*a*b* values between spherical geometry (C5-5) and 0/45 geometry (X-Rite 938). Accordingly, all visual observations were compared with spherical spectrophotometric measurements, specular included.

There was virtually no difference in visual observations made under D50 and Illuminant A light sources. To the question "Is there a color difference?" there were 305 yes votes under D50 light source, and 296 yes votes under Illuminant A. Consequently, all visual observations reported were for D50 light source.

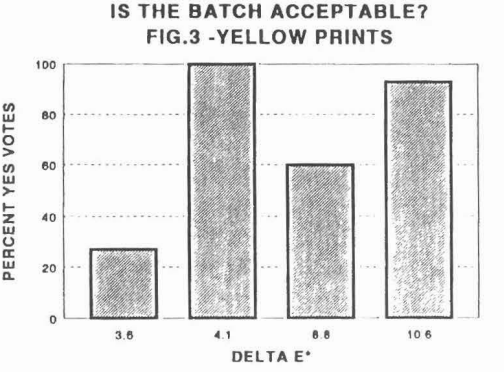
The panelists response to the questions "Is there a color difference?", and "Is the batch commercially acceptable?" is reported in Figures 1 through 7. Figure 1 shows the response for the hues orange, brown, and gray for which there were two samples each. There is good agreement with percent yes votes and DE* for the orange and brown samples. Note that 33 percent of the panelists observed a color difference for Brown 19 at 0.4DE* whereas 100 percent reported a difference for Brown 20 at 1.0 DE*. Gray is a very sensitive hue in that 87 and 80 percent reported a color difference at DE*'s of only 0.3 and 0.5, respectively. Orange is a less sensitive hue in that only 47 percent (7 panelists) reported a color difference when the batch and standard differed by 1.5 DE*.



Yellow samples are shown separately in Figure 2 because the DE* values for these samples were much higher than the other samples. There is an anomaly here in that panelists said unanimously that sample 26, which had the lowest (3.5) DE*, had a color difference whereas samples having higher DE's* did not have 100 percent ratings. Sample 26 was toned with red which resulted in an obvious color difference that was not apparent in the DE*value, however, there was a difference in Dh values.



Perhaps of greater significance in the evaluation of the yellow samples is the panel response to the question "Is the batch commercially acceptable?" Figure 3 shows that with a DE* value of 4.1,100 percent of the panel said the batch was acceptable, and even at DE's* of 8.8 and 10.7 the majority of the panel accepted the batch.



The red samples (Figure 4) were characterized by having a narrow and low range of DE* values. Nevertheless, there is a good correlation between DE* and color difference. The visual evaluation of red is very sensitive to DE* in that the majority of panelists saw color differences at DE's* of 0.4 to 0.64.

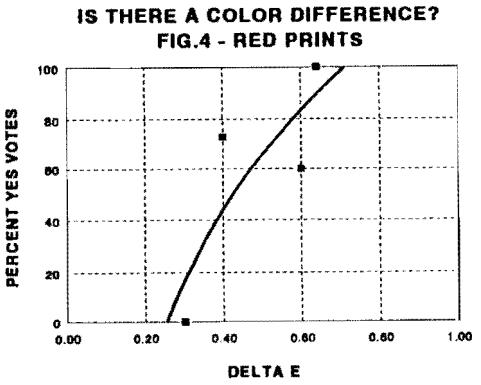
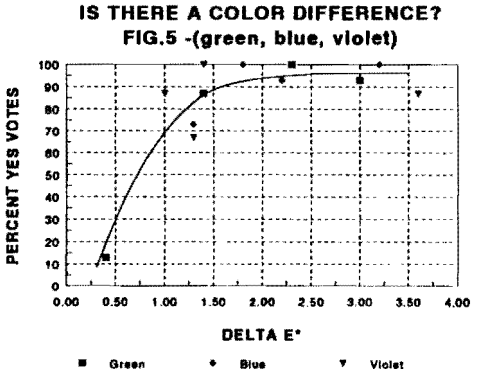
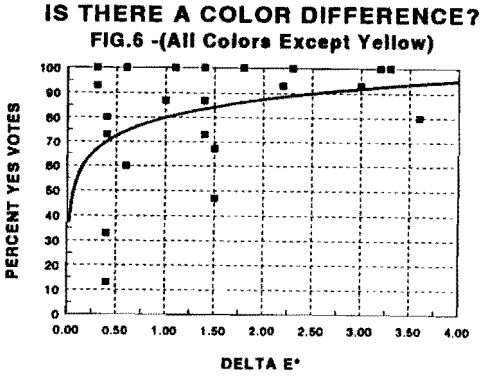


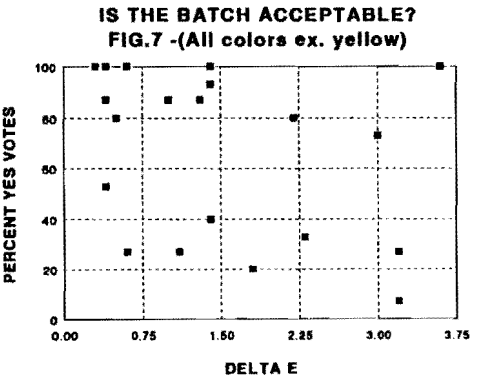
Figure 5 presents data for the violet, blue, and green samples which had similar DE* values and visual ratings. The correlation for each of these hues was good, with perhaps the best being green.



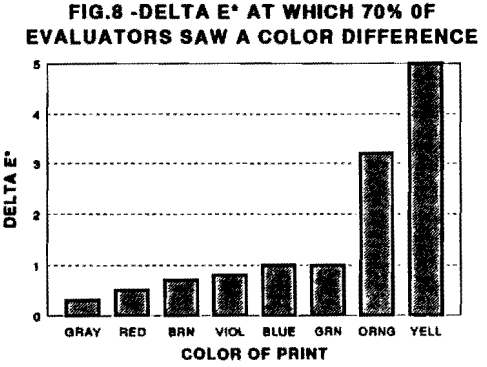


Color difference ratings for all samples, except yellow, are shown in Figure 6. As might be expected, there is considerably more scatter when seven hues (22 samples) are taken together. Although, there are data points in the upper left quadrant at low DE's* (i.e. a high percentage of color difference votes) there are none in the lower right quadrant (i.e. a low percentage of yes votes at high DE's).

Figure 7 presents the response to the question "Is the batch commercially acceptable?" for all samples except yellow. It is quite obvious that when a judgmental factor is introduced there is absolutely no correlation with DE*.

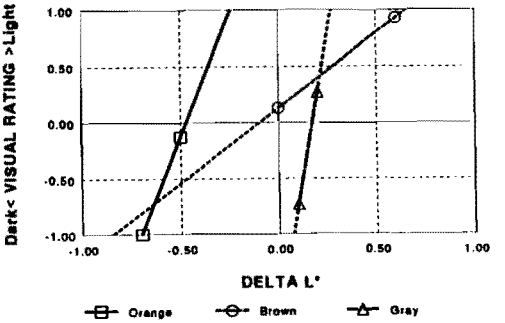


The DE* at which 70 percent of the panelists observed a color difference for each hue is shown in Figure 8. DE's* were determined by interpolation from data at higher and lower DE's* and visual responses. The DE's* where color differences are definitely observed varied from a low of 0.3 for gray to a high of 5.0 for yellow.

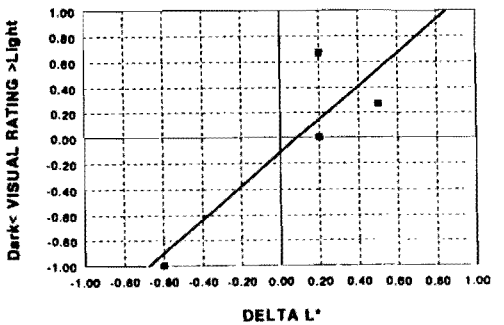


Lightness - The correlation between DL* and visual lightness ratings is shown in Figures 9-14. In general, the agreement between DL* over a range of -1.7 to + 0.9 and the visual lightness rating for all hues was very good.

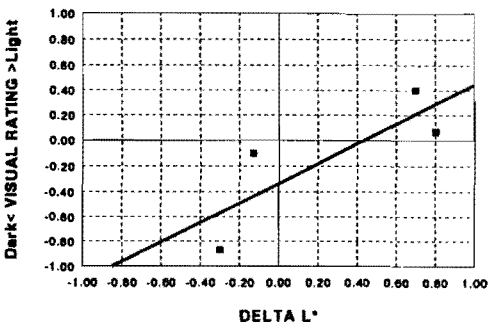
LIGHTNESS EVALUATION
FIG.9 - ORANGE,BROWN,GRAY



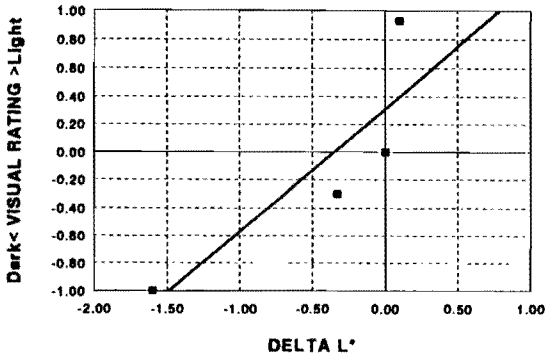
LIGHTNESS EVALUATION
FIG.10 -RED SAMPLES



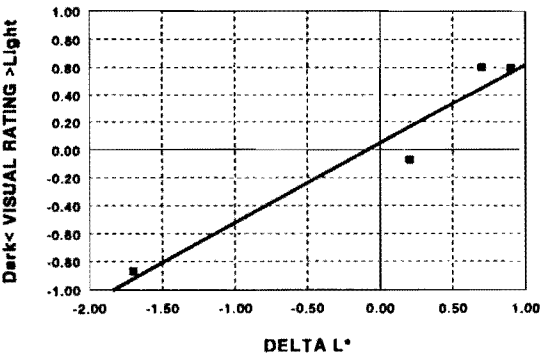
LIGHTNESS EVALUATION
FIG.11 -VIOLET PRINTS



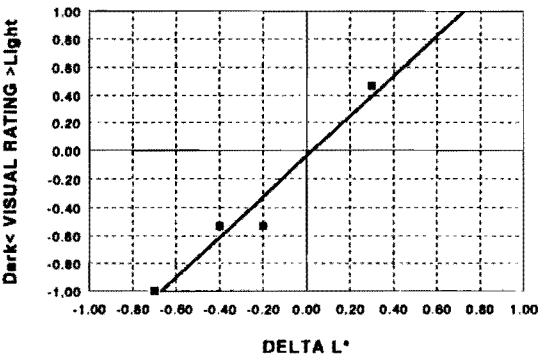
LIGHTNESS EVALUATION
FIG.12 -BLUE PRINTS



LIGHTNESS EVALUATION
FIG.13 -GREEN PRINTS

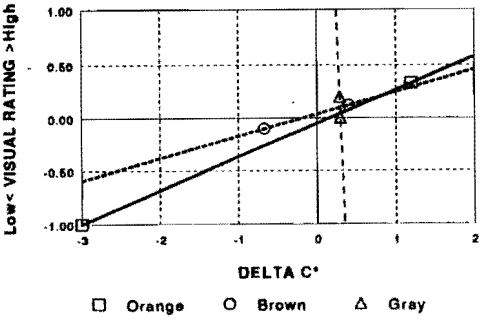


LIGHTNESS EVALUATION
FIG.14 -YELLOW PRINTS

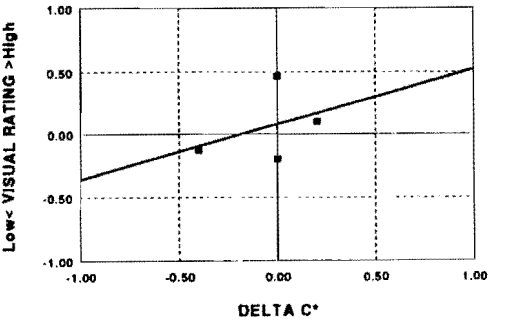


Chroma - Figures 15-20 present the visual rating for chroma for the eight hues over a DC* range of -4.0 to +10.5. The correlation between the visual rating and DC* was very good for all hues, particularly red, blue, green and yellow.

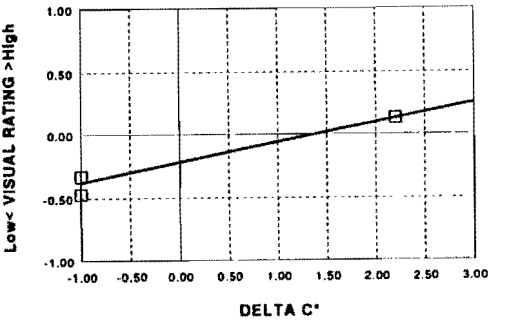
CHROMA EVALUATION
FIG.15 -ORANGE,BROWN,GRAY



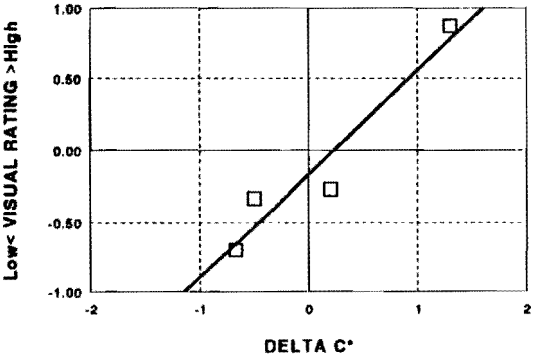
CHROMA EVALUATION
FIG.16 -RED PRINTS



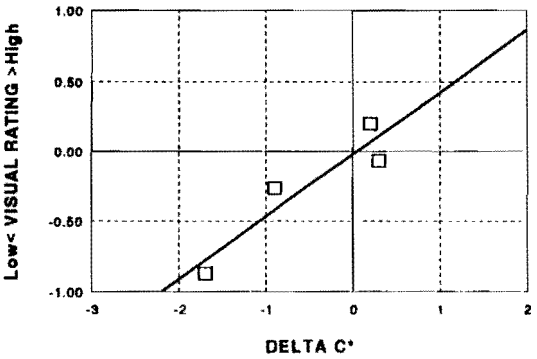
CHROMA EVALUATION
FIG.17 -VIOLET PRINTS



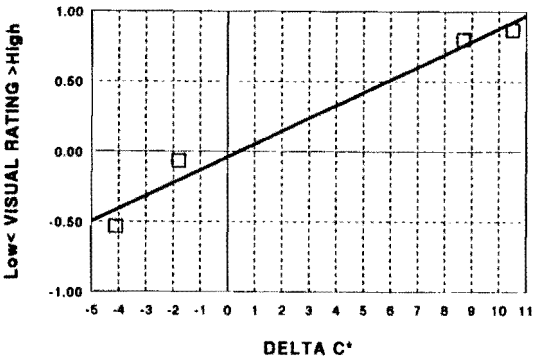
CHROMA EVALUATION
FIG.18 -BLUE PRINTS



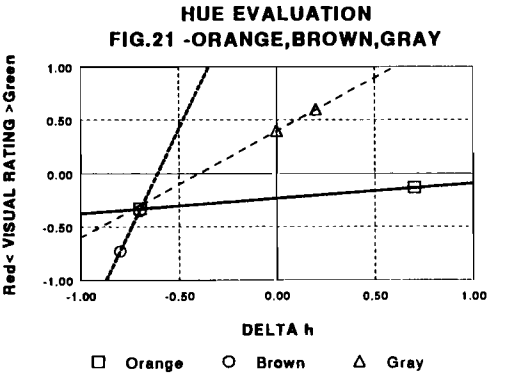
CHROMA EVALUATION
FIG.19 -GREEN PRINTS

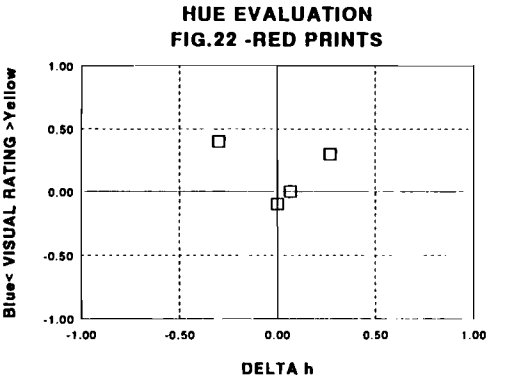


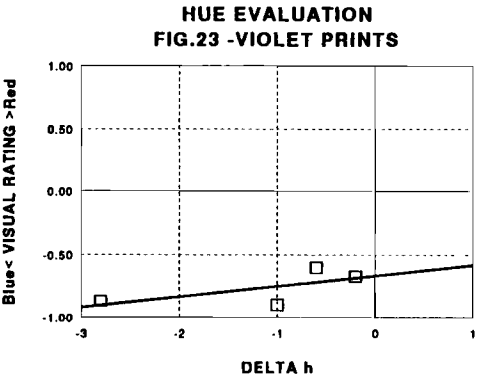
CHROMA EVALUATION
FIG.20 -YELLOW PRINTS



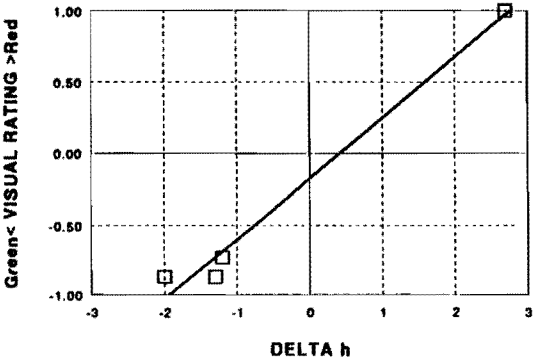
Hue - As shown in Figures 21-26, there was more scatter in the data when the visual rating for hue is plotted against Dh, with the exception of green (Figure 25) which was quite good over a Dh range of-1.2 to +2.3. Nevertheless, the overall qualitative agreement between Dh values and the visual ratings was good.



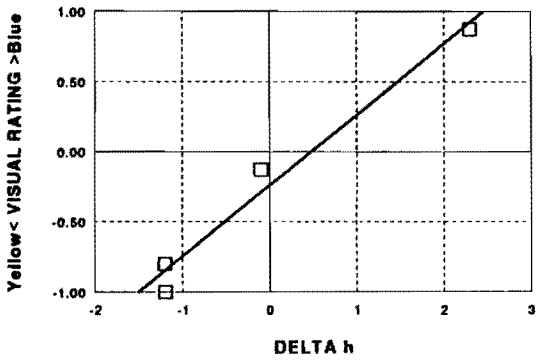




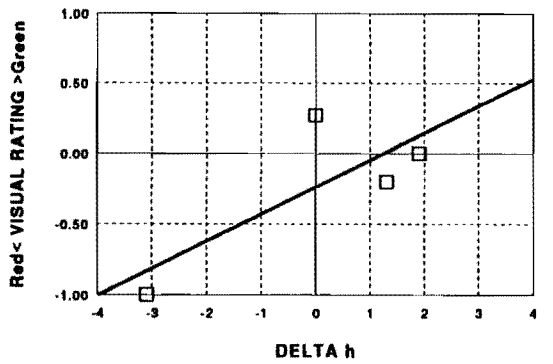
HUE EVALUATION
FIG.24 -BLUE PRINTS



HUE EVALUATION
FIG.25 -GREEN PRINTS

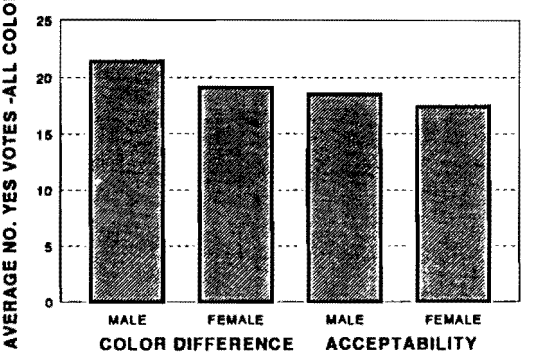


HUE EVALUATION
FIG.26 -YELLOW PRINTS



Gender Comparisons - A comparison of male/female voting patterns is shown in Figures 27-30. For the questions of color difference and batch acceptance (Figure 27) there is an interesting reversal. The average number of yes votes per male panelist to the question "Is there a color difference?" was 21.4 compared with 19.1 for the female panelists, an 11 percent difference. However, to the question "Is the batch commercially acceptable?", the average number of yes votes was 18.5 for the men, and 17.4 for the women.

FIG.27 -GENDER COMPARISON
COLOR DIFFERENCE & ACCEPTABILITY



In rating lightness (Figure 28), there was virtually no difference between the male and female averages for the ratings "darker", and "no difference". On the other hand, the male average for lightness votes was 7.5 compared with 5.6 for the female average, a 25 percent difference.

FIG.28 -GENDER COMPARISON
LIGHTNESS OF ALL PRINTS

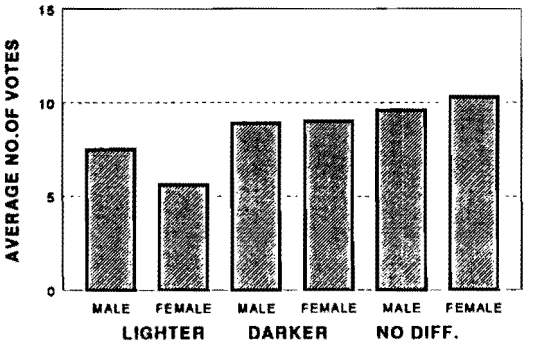
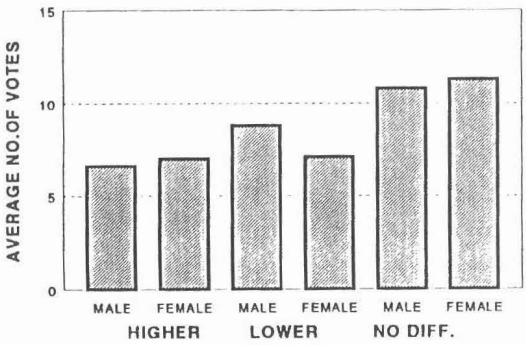


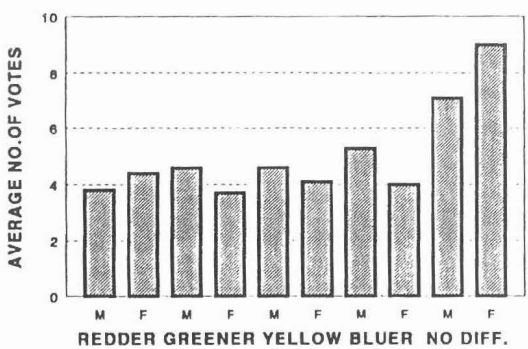
Figure 29 shows a difference of 19 percent (8.8 vs. 7.1) in the male/female rating of "lower chroma". There was not much difference in the male/female voting pattern for higher chroma, and the "no difference" rating.

FIG.29 -GENDER COMPARISON
CHROMA OF ALL PRINTS



The male-female comparison for hue is shown in Figure 30. Slight differences were observed for the ratings "redder", "greener", and "bluer". However, the greatest difference occurred with the "no difference" rating where the average for men was 7.1 and for women, 9.0, a 21 percent difference.

FIG.30 -GENDER COMPARISON
HUE OF ALL PRINTS



A closer examination of male/female voting patterns is shown in Table 1.

Table 1

Male/Female Comparison, Yes Votes

	Color Difference		Batch Acceptance	
	Male	Female	Male	Female
Maximum	24	24	26	24
Mean	21.5	19.1	18.1	16.9
Minimum	17	14	12	10
Range	7	10	14	14
Std. Dev.	2.2	4.2	4.0	4.2

The point that should be emphasized here is that there was a big difference in ratings by individual panelists. Two female and one male panelist thought that 24 of the 26 color samples had a color difference, compared with only 17 and 14, respectively, for two other panelists. Turning to batch acceptance, one male panelist thought that all 26 batches were acceptable compared with only 12 by another male panelist. On the female side, one colorist thought that 24 of the 26 batches were acceptable compared with only 10 for another female panelist. The range in the number of yes votes was greater for the female panelists in judging color difference, which is also reflected in standard deviation, a measure of variability.

Discussion of Results

Visual Ratings - On an overall basis, the agreement between visual ratings and instrumental measurements was good. The degree of agreement appears to be related to hue. For some hues, and most color attributes, the correlation was almost quantitative. For other hues, the correlation was qualitative, at best.

Perceptible Color Difference - An important finding in this study was the DE* at which a majority (70 percent) of the panel observed a color difference. It perhaps is generally known that the observation of a color difference varies with hue. This study provides definitive information. For instance, the DE* at which a color difference was observed for gray, red

and brown was significantly less than 1.0. The DE* for violet, blue and green was around 1.0. and for orange and yellow it was greater than 2.2.

Male/Female Comparison - There were no dramatic differences observed when the average from the seven female colorists are compared with the average for the eight male colorists. However, there were some differences. In evaluating color difference, a slightly greater (11 percent) number of male colorists saw a color difference than female colorists. However, we do not believe that on the basis of this one limited study, we can say that male colorists are more critical, or have a greater perception of color than females.

The question "Is the batch commercially acceptable?" is, of course, important to ink manufacturers and their customers. This question is much more judgmental than the question about color difference, and the results bear this out. There was virtually no correlation with DE* value. Also, although men were in a sense more critical in judging color difference, they were slightly more forgiving or tolerant in judging batch acceptance.

Looking at the color attributes of lightness, chroma, and hue, a greater number of male panelists rated the batch lighter than female panelists by 25 percent. Although this is a significant difference we do not believe that a conclusion can be drawn based upon this one study. For chroma, the lower rating for male panelists was 19 percent greater than female panelists. This also is a significant difference, but again we do not believe conclusions can be drawn until studies like this are repeated.

Hue is a difficult color attribute to judge. Our results show some slight male/female differences that are not at all conclusive. There was, however, a larger difference in the average number of male colorists who voted "no difference" (7.1) compared with female colorists (9.0), a difference of 21 percent.

Individual Panelists - In reviewing the ratings submitted by each of the panelists, it was apparent that each panelist was fairly consistent in the way he or she voted. However, it was also apparent that some panelists were consistently more critical in their assessment of color differences, and others were consistently less critical in assessing the commercial acceptability of a batch. In fact, one male panelist did not reject any of the 26 batch samples, versus an average of 6 rejections for the entire panel.

Conclusions

1. For eight sets of colors covering a broad spectrum range the DE* at which a color difference was definitely observed varied from 0.3 for gray to 5.0 for yellow.

2. There was good agreement between average visual ratings and the color attributes DL*, DC* and Dh.

3. There was no significant difference in visual ratings when D50 and Illuminant A light sources were used. This is an indication that the samples selected had minimum metamerism.

4. The perception of color difference varied significantly between individual panelists.

5. Some differences were observed between average male and female visual ratings. The significance of these differences requires further study.

Acknowledgement

The Task Force wishes to acknowledge the efforts of Lara Bottone, J.M. Huber Company, who prepared and measured the print samples, and Nick Lena, Macbeth, who made spectral radiometer readings of the light sources.

APPENDIX TABLE 1

Sample Order of Evaluation

Sample Number	Hue
8	Violet
6	Red
16	Green
18	Green
25	Yellow
1	Orange
7	Violet
13	Blue
3	Red
10	Violet
26	Yellow
14	Blue
24	Yellow
19	Brown
4	Red
5	Red
22	Yellow
2	Orange
9	Violet
11	Blue
15	Green
23	Yellow
12	Blue
21	Gray
20	Brown
17	Green

APPENDIX TABLE 2

NPIRI TASK FORCE ON COLOR MEASUREMENT

COLOR EVALUATION FORM

SAMPLE #: _____ HUE: _____ LIGHT SOURCE D50: _____ ILUM A: _____

1. Is there a color difference between the batch (B) and standard (S)? Yes _____ No _____
2. If (1) is "NO", go to item #14 and proceed to next sample.
3. If (1) is "YES", is the batch lighter or darker? _____ No difference _____
4. If (1) is "YES", is the chroma of the batch more saturated (cleaner, higher chroma) or less saturated (dirtier, lower chroma)? _____ No difference _____

For the applicable hue, please answer one of the following questions about the hue, then proceed to item #13:

5. For ***orange***, is the batch redder or greener? _____ No difference _____
6. For ***red***, is the batch yellower or bluer? _____ No difference _____
7. For ***violet***, is the batch redder or bluer? _____ No difference _____
8. For ***blue***, is the batch greener or redder? _____ No difference _____
9. For ***green***, is the batch yellower or bluer? _____ No difference _____
10. For ***brown***, is the batch redder, greener, yellower, or bluer? _____ No difference _____
11. For ***grey***, is the batch redder, greener, yellower or bluer? _____ No difference _____
12. For ***yellow***, is the batch redder or greener? _____ No difference _____
13. Is the batch (B) acceptable as a commercial match? Yes _____ No _____
14. _____ _____ _____
    Signed **Company** **Date**

Age: ___(< - 30), ___(30 - 50), ___(> - 50) Gender: M _____ F _____ Years Experience: _____

D50/2 degree
CIELAB

	0/45 X-Rite	D/8 CS-5	0/45 X-Rite	D/8 CS-5	0/45 X-Rite	D/8 CS-5	0/45 X-Rite	D/8 CS-5	0/45 X-Rite	D/8 CS-5	0/45 X-Rite	D/8 CS-5
Sample	DL*	DL*	Da*	Da*	Db*	Db*	DC*	DC*	Dh	Dh	DE*	DE*
1 ORANGE	-0.71	-0.71	-1.74	-1.85	-2.54	-2.51	-2.98	-3.04	-0.57	-0.68	3.15	3.19
2 ORANGE	0.64	0.51	0.78	0.44	1.63	1.35	1.66	1.22	0.50	0.72	1.91	1.51
3 RED	0.18	0.17	-0.05	-0.10	0.01	-0.18	-0.40	-0.19	0.03	-0.10	0.19	0.27
4 RED	0.61	0.47	-0.39	-0.25	-0.58	-0.28	-0.64	-0.36	-0.19	-0.09	0.93	0.60
5 RED	-0.56	-0.55	0.22	-0.05	0.75	0.33	0.48	0.08	0.54	0.32	0.97	0.64
6 RED	0.07	0.20	0.18	0.08	-0.24	-0.33	0.08	-0.04	-0.26	-0.34	0.31	0.40
7 VIOLET	1.09	0.77	-1.05	-1.06	-0.44	-0.43	-0.98	-0.99	-0.68	-0.57	1.57	1.38
8 VIOLET	0.03	-0.06	-0.89	-0.97	-0.11	-0.03	-0.86	-0.95	-0.28	-0.17	0.89	0.97
9 VIOLET	0.90	0.70	0.52	0.62	-3.42	-3.44	2.02	2.16	-2.58	-2.75	3.57	3.57
10 VIOLET	0.03	-0.30	-1.78	-1.24	-0.40	-0.31	-1.36	-0.92	-1.10	-0.88	1.83	1.31
11 BLUE	-0.06	0.03	-0.31	-0.48	1.41	1.20	-0.80	-0.54	-1.75	-1.17	1.45	1.29
12 BLUE	-0.15	-0.32	-1.01	-0.93	2.01	1.92	-0.72	-0.73	-3.10	-2.00	2.25	2.16
13 BLUE	-1.13	-1.59	2.27	2.74	0.41	0.13	-0.14	0.17	2.75	2.74	2.57	3.17
14 BLUE	0.23	0.13	-1.41	-1.22	-1.32	-1.35	1.26	1.30	-1.92	-1.28	1.95	1.83
15 GREEN	0.11	0.18	-0.19	-0.25	0.10	0.18	0.22	0.30	-0.04	-0.08	0.25	0.36
16 GREEN	0.84	0.87	2.19	2.02	0.57	0.53	-1.84	-1.69	-2.47	-1.22	2.41	2.26
17 GREEN	0.37	0.65	0.92	0.65	0.60	1.01	-0.17	0.20	-0.94	-1.18	1.15	1.37
18 GREEN	-1.30	-1.65	-0.59	-0.89	-2.80	-2.31	-1.44	-0.89	2.19	2.31	3.15	2.97
19 BROWN	-0.16	0.02	0.00	0.13	-0.38	-0.33	-0.18	-0.07	-1.05	-0.35	0.41	0.36
20 BROWN	1.07	0.55	0.89	0.76	-0.36	-0.50	0.56	0.35	-3.11	-0.84	1.43	1.06
21 GRAY	0.33	0.08	-0.17	-0.12	0.53	0.28	0.54	0.30	1.52	0.04	0.65	0.32
22 GRAY	0.33	0.20	-0.26	-0.22	0.26	0.24	0.30	0.29	5.90	0.15	0.49	0.38
23 YELLOW	0.32	0.34	-0.20	-0.12	-4.55	-4.11	-4.54	-4.11	0.04	0.04	4.57	4.13
24 YELLOW	-0.07	-0.22	-0.08	-0.36	11.87	10.65	11.70	10.47	1.33	1.94	11.87	10.65
25 YELLOW	-0.03	-0.38	-0.18	0.03	8.86	8.81	8.72	8.72	1.12	1.27	8.87	8.82
26 YELLOW	-0.80	-0.72	2.62	2.48	-2.36	-2.55	-1.56	-1.79	-2.10	-3.07	3.61	3.63

APPENDIX TABLE 3