REPORT

on Colour Fidelity and Colour Preference investigations

Executive Summary

Colour fidelity, i.e. visual colour difference evaluation, using Munsell colour samples and colour preference investigations, using both ISO test images and real fruits and vegetables, were carried out by 19 naïve observers and six trained colour observers ("experts"). They compared the test samples under incandescent illumination with three "modern" light sources, a CFL, a commercial white LED lamp and a Pharox Bulb by Lemnis Lighting with light-engine (XED module) by Ledon (PL/LED bulb) (in the following P/L LED-bulb). In colour preference the three fundamental perceptions: naturalness, vividness and preference were tested.

Preference experiments have shown superiority of the P/L LED-bulb. Most pronounced was this superiority in "naturalness".

In colour fidelity the visual colour difference between the colour of the Munsell samples illuminated by the incandescent reference lamp (in the following Inc) and the test lamps has been evaluated. Experiments have confirmed that colour difference – if evaluated in an up to date colour space – describes colour fidelity quite well.

Introduction

The standard method of colour rendering calculation[1] describes colour rendering of a light source by determining calculated colour differences in the CIE U*,V* W* space, using 8 + 6 colour samples, comparing their calculated colour difference when illuminated by the test source and a reference source of equal correlated colour temperature. This method of colour rendering calculation, where it is normal to average the values for the first 8 samples of lower chroma to reach to the General Colour Rendering index (R_a) was developed to distinguish between the colour reproduction under traditional halophosphate fluorescent lamps and the so called "deluxe" lamps with improved colour rendering.

Since the invention of the three-band fluorescent lamps observations have been aired that the calculated general colour rendering index does not describes the visual impression correctly. CIE has dealt with the problem in several technical committees and finally came to the conclusion that it is not enough to specify one colour rendering index, but that there are tasks, where colour fidelity is of high importance, and in other tasks colour preference is more essential, see CIE TC 1-69: "Colour rendition by white light sources"[2]. At present this TC is collecting visual data and it is hoped that within a reasonable time CIE will be able to come up with recommendations both for colour fidelity and colour preference indices.

In this report we will provide results of three experiments: One experiment was conducted to evaluate colour fidelity and two experiments for colour preference. Colour fidelity was tested by checking visual colour difference for ten Munsell samples corresponding well to the samples suggested by CIE TC 1-33 for colour rendering investigations[3], comparing the visual colour differences of the same samples observed once illuminated by an incandescent lamp and then by different test lamps. Colour preference was evaluated by comparing equal samples of four ISO test pictures seen under the Inc lamp and the test lamps, asking observers whether under the test light source the picture was more vivid, looked more natural and whether it was preferred or not. As a second experiment for preference a group of ten items (fruits and vegetables) were presented in every booth, and the observers had to report for each item whether the colour of the given object was more vivid, looked more natural or more preferred. This second preference experiment was an absolute scale experiment, where the observers compared the colour of the seen objects with his/her mental colour image of the given object type.

19 naïve young observers, who had not much experience with colour scaling and six "experts", persons who deal with colours in their everyday tasks, took part in the experiments. We tried to use the two groups to test whether a difference in preference could be found or not.

Test equipment and samples

Light sources

The following samples were analysed:

- Osram 40 W h 938 incandescent lamp (Inc)
- Osram DuluxStar Classic A 7W/827 Lumilux WW (CFL)
- Philips Master LED BULB MV Dimmable, 7W 3000 K (Ph-LED)
- Pharox Bulb by Lemnis Lighting with light-engine (XED module) by Ledon (P/L LED bulb)
- RGB-LED cluster, with emission maxima at 626 nm, 525 nm and 455 nm

Test booths

Five booths have been constructed, where in the upper compartment place was available for the light source, below of which an opal acrylic plate was placed to diffuse the light properly. This upper compartment was painted white. The lower compartment, below the acrylic plate was painted grey, with an L^* value of 70,6, and a^*=-0,72, b^*=-1,24.

By changing the distance of the light source from the acrylic plate the illuminance at the bottom of the five booths was set to an approximately equal value, as seen in Table 1.

Table 1. Illuminance in the single booth

Number of light source/booth	Light source	Illuminance, lux
L-00	IncW	721
L-01	RGB-LED	716
L-02	CFL	684
L-03	Ph-LED	682
L-04	P/L LED-bulb	690

Figure 1 shows the five booths, with the black blinds lifted.

Figure 1. Picture of the five test booths.

Test samples

Three sets of test samples have been used: Munsell chips, ISO pictures and real fruits and vegetables.

Munsell samples

From a Munsell sample collection chips have been selected with the nearest Munsell notation to the Munsell notations of the samples used in the Macbeth Color Checker Chart (MCCC). Table 2 shows the ISCC/NBS name of the samples, their CIE number in[1], the MCCC number, the Munsell notation and the measured CIELAB co-ordinates.

Table 2. Munsell samples used in the experiment

ISCC/NBS Name	CIE TCS Sample #	MCCC Sample #	Munsell notation	*L**	*a**	*b**
strong red	1	1	5R4/12	40.91	53.32	24.73
strong orange	2	12	5YR6/10	60.73	28.99	49.60
vivid yellow	3	9	5Y8/12	81.29	-1.68	82.78
strong yellow green	4	7	5GY7/8	71.44	-24.25	49.17
strong yellowish green	5	10	2.5G5/10	51.94	-50.89	25.02
strong greenish blue	6	2	5B5/8	51.49	-24.69	-25.13
vivid purplish blue	7	5	7.5PB3/12	31.54	20.15	-46.54
strong reddish purple	8	6	2.5RP5/12	50.93	48.93	-16.29
light reddish brown	9	4	2.5YR6/4	61.47	14.03	17.04
moderate brown	10	3	2.5YR4/4	41.30	16.17	17.92

ISO test pictures

The following four pictures have been selected from the ISO Set for graphic arts quality evaluation[4]: N1 Young lady with the Glass, N2 Flowers, N6 Pier and N7 Threads.

Figure 2.:N1 Young lady.

Figure 3. N2 Flowers.

Figure 4. N6 Pier.

Figure 5. N7 Threads.

These pictures enable the evaluation of skin tones, colours of flowers, a landscape image and vivid everyday colours. From every image two identical copies were printed for pair comparison in the reference booth and each of the test booths.

Fruits and vegetables

Real fruits and vegetables served for absolute evaluation of the different illuminations. Ten well known objects have been selected and samples have been spectrophotometrically measured before and after every usage. Reproducibility of these spectrophotometric measurements was 3 to 4 ΔE_{ab} units. If the colour of an object changed more than 5 units it was replaced by a fresh sample. Practically every other day new samples of the banana, the different paprika samples and the tomato had to be used.

Objects have been selected to be of well known colour also for naïve observers, and should cover the critical area of the colour space, mainly the reds and yellows. Blue colours are less critical for colour evaluation, and are well covered in the photos and Munsell chips. For the time period of the year, when the experiments were carried out, no blue fruits were available, where one would have been able to state a mental (memory) colour.

Table 3. CIELAB values of fruits and vegetables, Illum. A, 2° observer

	*L**	*a**	*b**
1 banana	82,22	7,54	53,66
2 grapefruit	85,41	11,83	65,42
3 orange	74,98	34,36	80,55
4 lemon	83,00	11,19	71,86
5 carrot	63,68	32,05	42,20
6 tomato	46,54	43,10	37,17
7 green paprika	40,29	-12,44	19,72
8 yellow paprika	78,26	1,94	43,42
9 red paprika	37,97	40,60	23,01
10 green leaf	38,19	-13,83	20,06

Experimental procedure

Introducing the observers to the task

As a first step the colour vision of the observers was tested using a Hungarian variant of the Ishihara test[5], only persons with normal trichromatic vision have been selected as observers. Naïve observers got a short introduction to the three-dimensional description of a colour perception, and using the NCS tutorial samples they had to order some coloured chips according to hue, lightness and chroma.

As a next step observers got an explanation of the three tasks to perform:

- Using the *memory colours* of the fruits and vegetables observers were asked to evaluate for vividness, preference and naturalness on a -5 to +5 scale the colour of the ten objects seen in one of the booths at a time (the other booths were blanked by black blinds). (Thus e.g. if a fruit colour was found to be un-natural in one booth it could have been marked in the evaluation sheet as -3.). In this open scale experiment observers were asked to put at the end of the test their hands into the booths and evaluate also the skin tone according to the three criteria. In this study for all five booths observers gave marks for the three criteria for each of the 10 objects and their hands.
- *Pair comparison*: One copy of the four pictures introduced in section *ISO test pictures* was put into the booth lit by the Inc. lamp. It was explained to the observers that this is the picture now seen under a reference illuminant. A second copy of the picture was put in one of the test booths, and the observers were asked to compare the two pictures and tell whether they found the picture in the test booth more appealing (preference), more vivid, more natural then under the reference source (Inc.). Only the two booths (ref. and just used test) were seen at this stage, the other booths were covered by the black blinds. The test image was successively put into all four test booths and the same questions answered using an evaluation sheet. Observers were instructed to use also here the -5 +5 scale.
- *Colour fidelity* experiment: The third task was to evaluate the colour difference of the

Munsell chips (introduced in section *Munsell samples*): First a chip was placed into the reference booth (illuminated by the Inc lamp), and subsequently into one of the test booths. Observers were asked to memorize the colour of the chip as seen in the reference booth, and subsequently when seen in the test booth assign a colour difference to the perceived colour. Naïve observers only scaled the colour difference on a 0 to 5 scale, expert observers were asked to evaluate also the three components of the colour change: change in hue, in chroma and lightness. This was repeated with every colour chip.

In the colour preference experiments the three variables were naturalness, preference and vividness.

- Naturalness is the simplest variable: in the memory colour experiment the observers had to answer whether the object looked similar to their mental image of the given object, in the pair comparison they were told that the image in the booth illuminated by the incandescent lamp is the reference, the master of the image wanted to show those colours.
- Preference is also an easy variable: in both cases the question was: would the observer like to see the object/picture under the given lamp, in the first case (memory colour), even if the object looks not natural, but it looked nicer as the natural object they remember. For the images: would they prefer to see the just seen image under such a light in their room, compared to the reference illumination.
- Vividness is the most difficult variable, especially for the objects (memory colours), as an object might look more vivid, i.e. more highly coloured (colourfulness?), but this might be unnatural at the same time. In case of the pictures one could formulate it somehow as if you look at such pictures you feel cheerful or sad.

Performing the experiment

Lamps were lit at least 30 minutes before starting visual observation. The level of illumination was checked periodically and was found to stay in a +/-5 % range. As the final aim of the experiment was to see how modern light sources acceptance compares to that of incandescent lamps, a rigid sequence was used_ Inc., RGB-LED, CFL, Ph-LED, P/L LED-bulb, but the observers did not know in which booth which lamp illuminates the samples. (Showing first the Inc. and the RGB-LED enabled the observers to compare to these two observations the other samples.)

As the *memory colour scaling* (absolute scaling) of the objects was regarded as the simplest task, usually observers started with that test. The most often used arrangement of the 10 objects is seen in Figure 6.

Figure 6. Arrangement of objects in the memory colour experiment.

The *pair-comparison* experiment was used as second. Here always two booths were uncovered, the reference one and the just under investigation being test booth. Observers tested one image after the other, placing one copy of the image into the reference (Inc.) booth, and the other in the uncovered test booth. Scaling for all three attributes (vividness, preference and naturalness) was made for each picture pair. Observers were allowed to re-check a picture in the reference booth several times before they made their judgment for one of the attributes, but were asked go closely to the booth they were just looking at, and wait for a few moments to adapt to the light of the given booth (see colour temperature difference between the booths in Table 4, section ***Results,*** *objective criteria*). In this experiment the attributes of the pictures in the reference booth was fixed as Zero, deviation from these had to be scaled.

The *colour fidelity* experiment using the Munsell colour chips for visual colour difference scaling turned out to be the most difficult, and it was left to the end of the session to get also the naïve observers accustomed to the scaling task. One of the ten Munsell chips was mounted on a grey pad, by the help of which the observer could insert the sample into one of the booths. In that case all booths were uncovered, and one expected a mixed adaptation of the observers, but with very little deviations, as both the CCT and the illuminance were very similar in all five booths.

Observers had to memorize the colour seen in the reference booth (Inc) and scale the difference when the chip was placed in one of the test booths. A grey scale was placed permanently in the reference booth, and observers were asked to use this scale to estimate the colour difference seen in the two booths. Naïve observers were unable to re-scale the observed absolute colour difference into three components: hue-, chroma- and lightness-difference, but expert observers gave also estimates of the three components. Observers were permitted to place and re-place the Munsell chip in one or the other booths, comparing all colour differences, but their estimate had to be against the reference booth colour.

Results

Objective criteria

The correlated colour temperature in each booth was determined by placing a PTFE white standard into the booths and measuring the reflected spectral radiance with a spectroradiometer.

Table 4. Measured CCT and CRI in the five booths

No of light source/booth	Light source	CCT, K	CRI, Ra
L-00	Incandescent	2660	99,1
L-01	RGB-LED	2664	41,3
L-02	CFL	2554	83,9
L-03	Ph-LED	2916	87,6
L-04	P/L LED-BULB	3037	90,7

For an eventual further experiment it would be good to select an incandescent lamp with higher CCT (low voltage halogen incandescent lamps exist that could be used), as the 300 to 400 K CCT difference might influence the observation, especially the vividness scaling.

Visual observations

In evaluating the visual observations first we had to realize that not every observer was able to scale properly on the -5 +5 scale. The total length of the scales used by different observers differed considerably, therefore we re-scaled the input data, by stretching/compressing the scale to get to comparable results. In the figures below average values and standard deviations are shown.

Memory colour scaling: The next figures show results of the absolute (memory) colour scaling. In these figures the scores given for the different objects (fruits and vegetables, hand) have been averaged, we show the results of the naïve, of the expert and of the entire group. The first three figures show results for vividness.

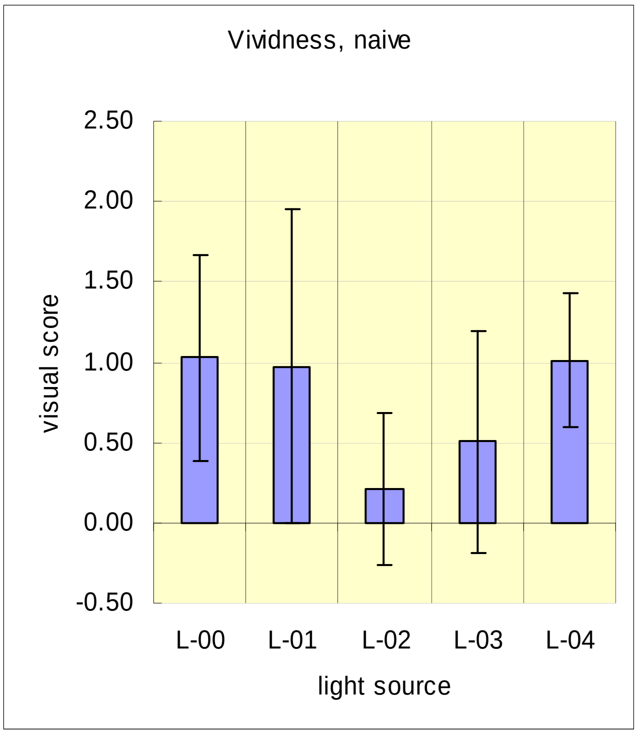


Figure 7. Memory colour scaling, vividness, naïve observers.

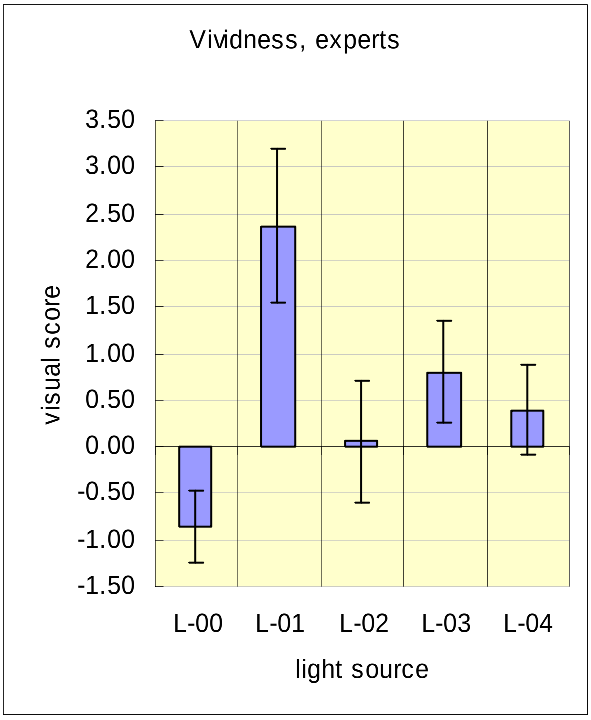


Figure 8. Memory colour scaling, vividness, expert observers.

Both groups rated the CFL low. Interesting is also that the two groups found the vividness of the real objects under L-03 and L-04 to be of opposite order. The main finding is that for real objects vividness is not necessarily an unambiguously positive feature.

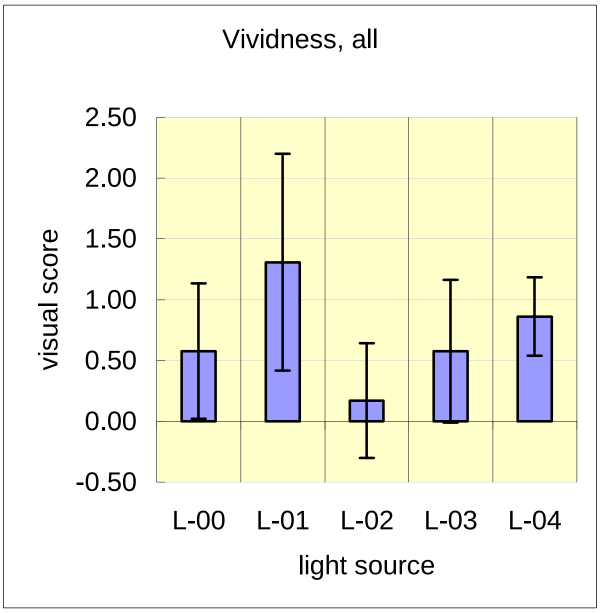


Figure 9. Memory colour scaling, vividness, all observers.

The next three pictures show results for preference, here the question was: "would you like have such lighting in your room or kitchen?"

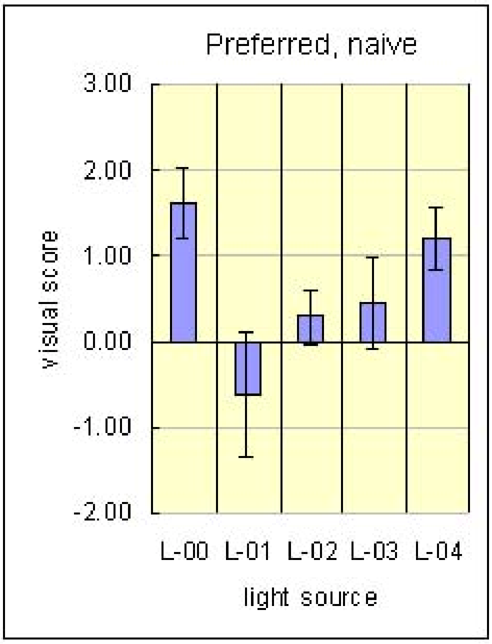


Figure 10. Memory colour scaling, preference, naïve observers.

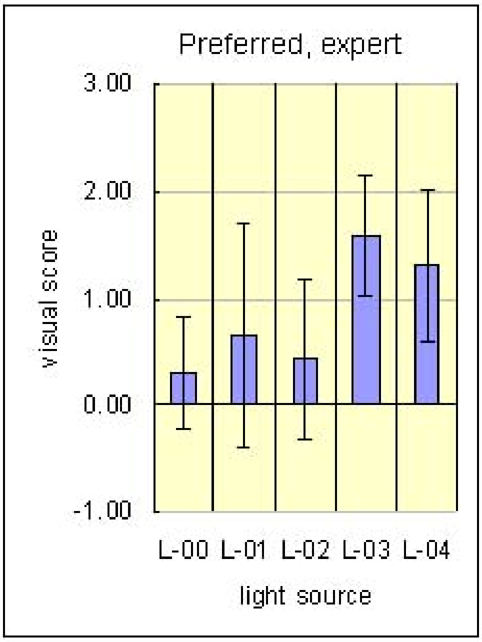


Figure 11. Memory colour scaling, preference, expert observers.

Difference between naïve and expert observers is seen here too. The naïve observers liked the traditional incandescent light, and both groups voted the CFL low. Interesting here is also the difference in judging the two white-LED lights.

Figure 12 shows the average scores of the entire group. The general over all viewpoint is obviously that the P/L LED-bulb provides an illumination that is equally preferred as incandescent light, all other lights score lower, the CFL is not a good replacement of an incandescent lamp; difference to Inc and P/L LED-bulb is insignificant.

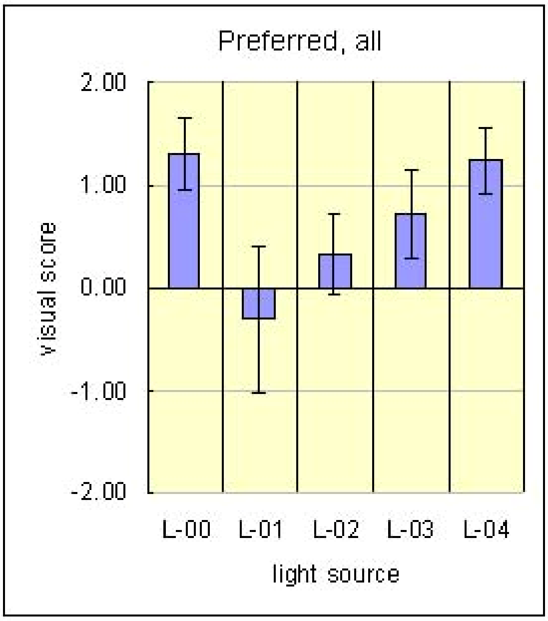


Figure 12. Memory colour scaling, preference, all observers.

The third question was on *naturalness*. The next three figures show the scores again of the naïve, the expert and the total group. The general trend of the answers to this question is the same for the two groups of observers; only the relative evaluation of the new white LED lights compared to the incandescent is different, probably due to the more conservative nature of the naïve observers.

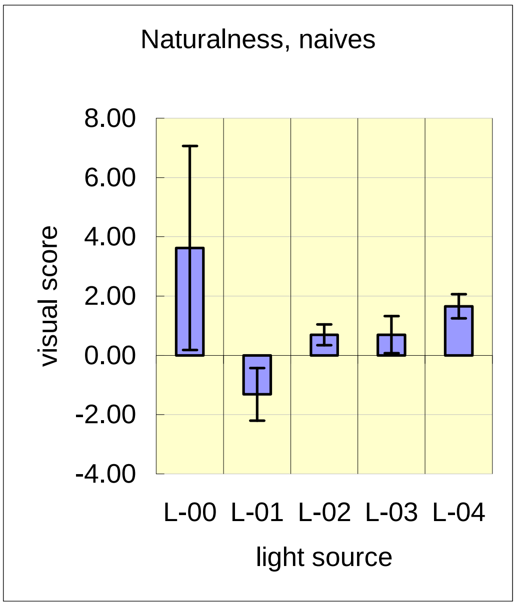


Figure 13. Memory colour scaling, naturalness, naïve observers.

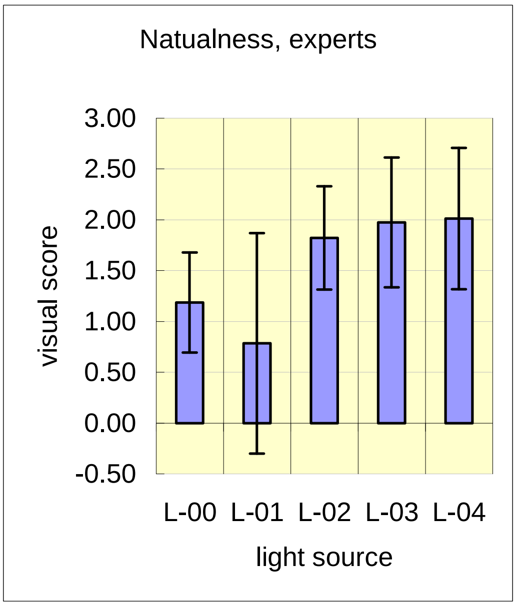


Figure 14. Memory colour scaling, naturalness, expert observers.

Figure 15 shows the summed up results of all the replies. The P/L LED-bulb scores equal to the incandescent and surpasses all other modern light sources. Interesting that for naturalness the CFL and the Ph-LED score equal.

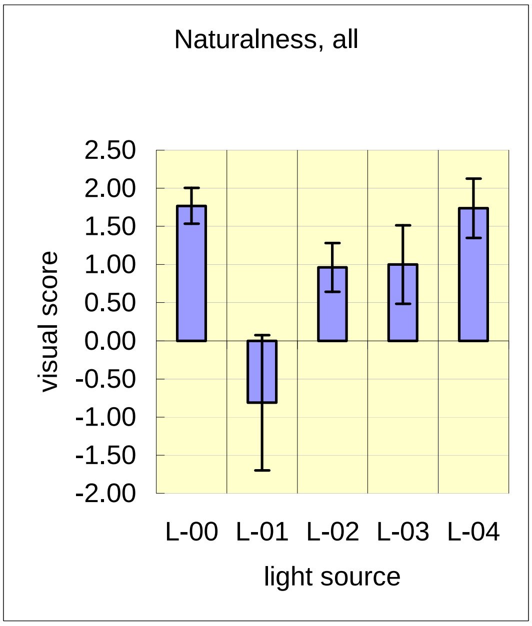


Figure 15. Memory colour scaling, naturalness, all observers.

As the number of lamps is relatively low it is an interesting question what the rank order of the different lamps is. The next figures show these results. Visual score "4" is the lamp that was chosen most frequently as the most "vivid", "preferred" or "natural". In this evaluation the RGB-LED has been excluded, as it is not a candidate for replacing the incandescent lamp.

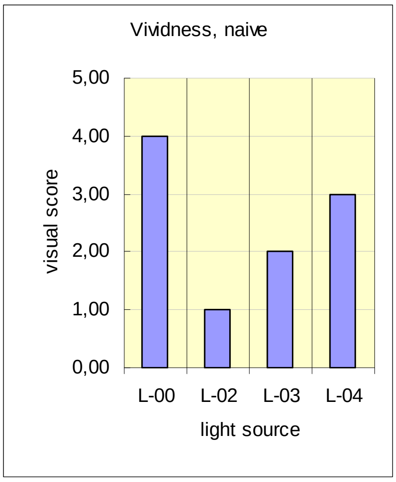


Figure 16. Rank order, memory colour scaling, vividness, naïve observers

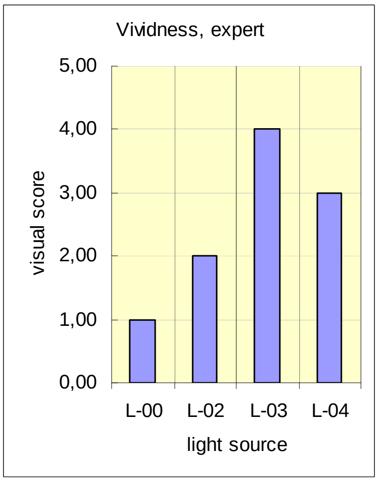


Figure 17. Rank order, Memory colour scaling, vividness, expert observers.

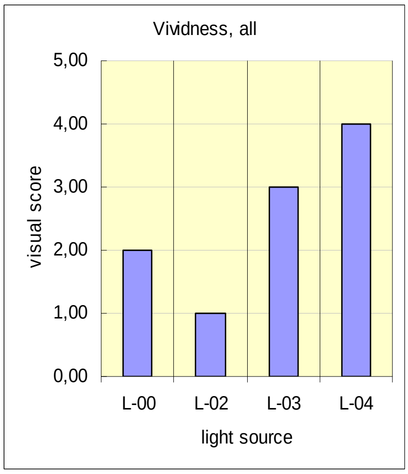


Figure 18. Rank order, memory colour scaling, vividness, all observers.

As can be seen, the CFL is rated low for *vividness*. The big difference in the naïve and expert rating of the incandescent lamp is still not understood, although this influences the rating difference between the two LED lamps. P/L LED-bulb rates highest for the combined group.

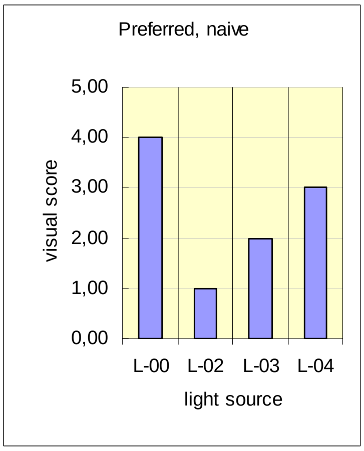


Figure 19. Rank order, memory colour scaling, preference, naïve observers.

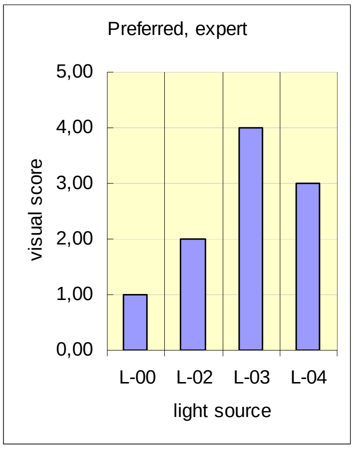


Figure 20. Rank order, memory colour scaling, preference, expert observers.

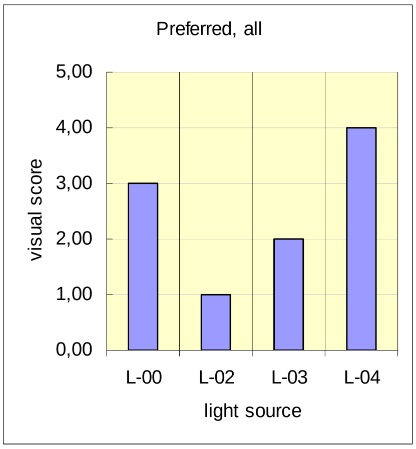


Figure 21. Rank order, memory colour scaling, preference, all observers.

As seen for *preference* the P/L LED-bulb is rated highest by the combined group, the scores are very similar to the "vividness" scores.

The evaluation for *naturalness* provides some further interesting results: The P/L LED-bulb stands for both the expert and combined groups on place No. 1.

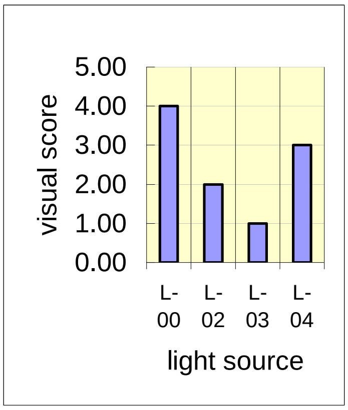


Figure 22. Rank order, memory colour scaling, naturalness, naïve observers.

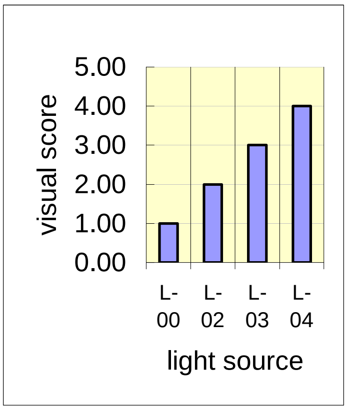


Figure 23. Rank order, memory colour scaling, naturalness, expert observers.

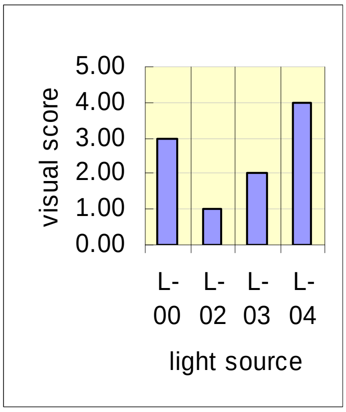


Figure 24. Rank order, memory colour scaling, naturalness, all observers.

Pair-comparison experiment

The pair-comparison experiment provided further information on the acceptance of a multicolour display, as seen in real world. The next three times three figures summarize the results of the pair-comparison experiment. Again the figures show the average scores for the four images.

Pair-comparison, vividness:

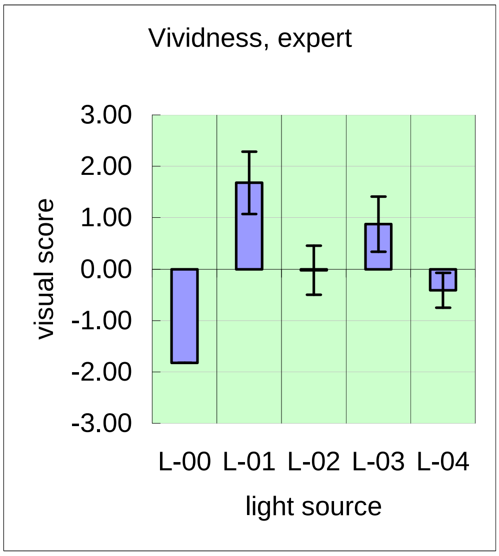


Figure 25. Pair-comparison, vividness, naïve observers.

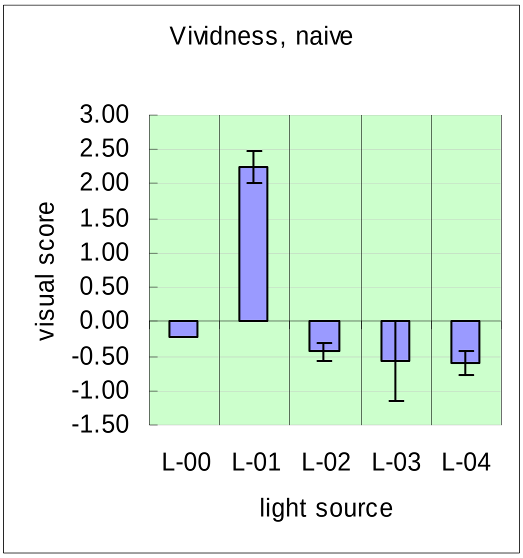


Figure 26. Pair-comparison, vividness, expert observers.

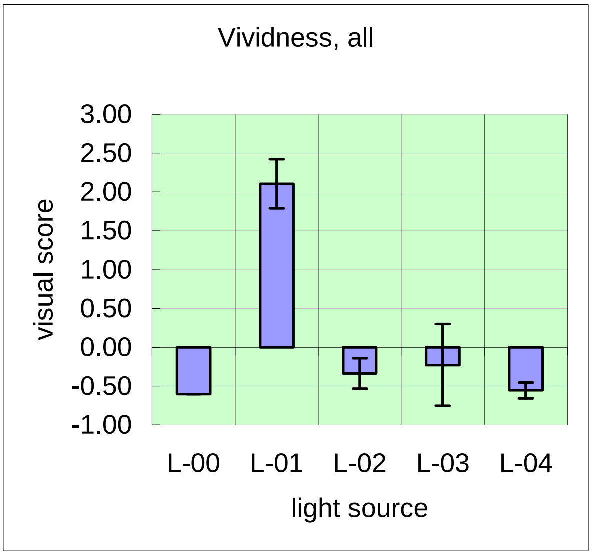


Figure 27. Pair-comparison, vividness, all observers.

Interesting is that the experts voted very negative for the incandescent lamp. Both groups found the RGB-LED producing most illumination (but other scores might be low)

Answers on *preference* compared to the incandescent illumination were also different among naïve and expert observers: While experts preferred CFL lighting (they found it also quite natural!) naïve observers preferred the RGB-LED most. (The non-zero score for the Inc booth comes from the re-scaling of the scores.)

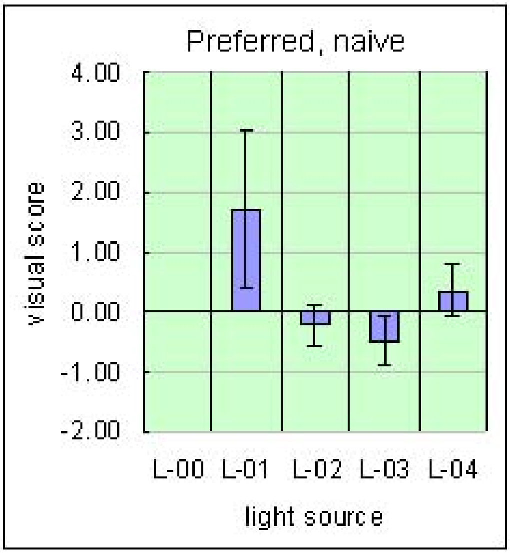


Figure 28. Pair-comparison, preference, naïve observers.

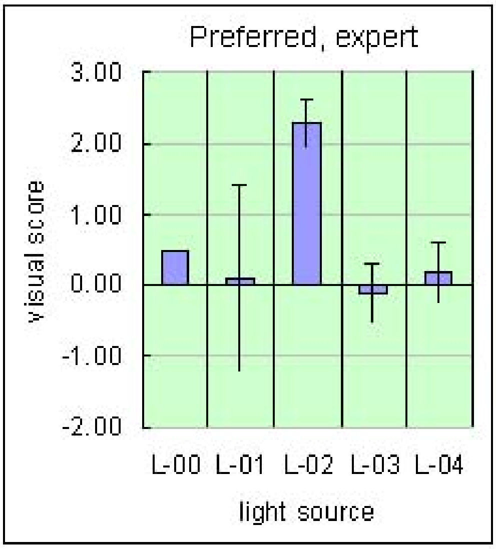


Figure 29. Pair-comparison, preference, expert observers.

Figure 30 shows the average of all observers. It can be seen that that the P/L LED-bulb scores equal to the CFL, an interesting contradiction to the scores for the real objects, where the CFL showed significantly lower preference as the P/L LED-bulb.

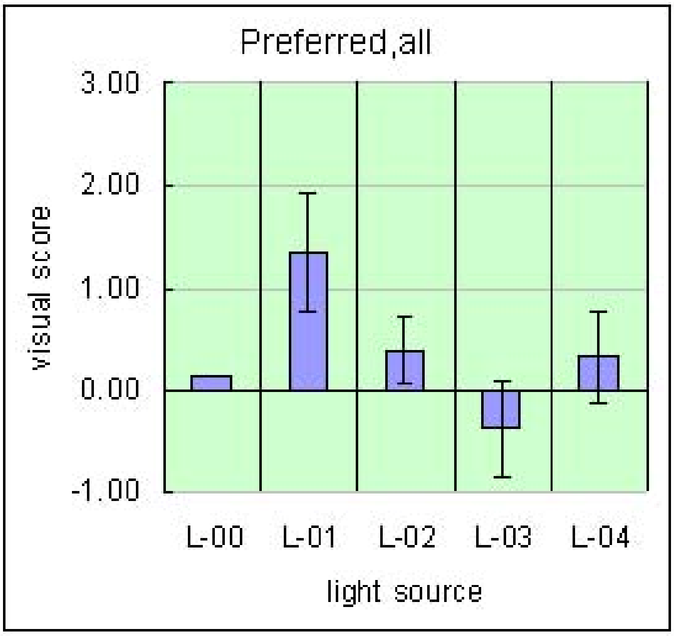


Figure 30. Pair-comparison, preference, all observers.

Naturalness is a critical quality parameter of the sources. The next three figures show the results: Also here the scores of the experts differed from that of the naïve observers. Experts were content with the CFL lighting, while naïve observers proved the general dislike of CFLs (they did not know which lamp was in which booths!).

In summary, see Figure 33, the P/L LED-bulb lighting was found most natural.

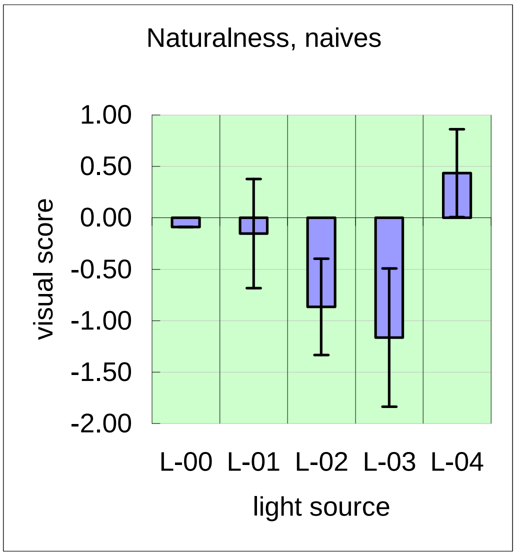


Figure 31. Pair-comparison, naturalness, naïve observers.

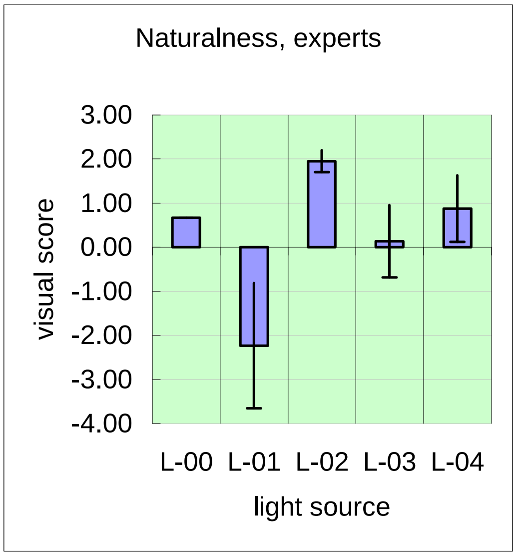


Figure 32. Pair-comparison, naturalness, expert observers.

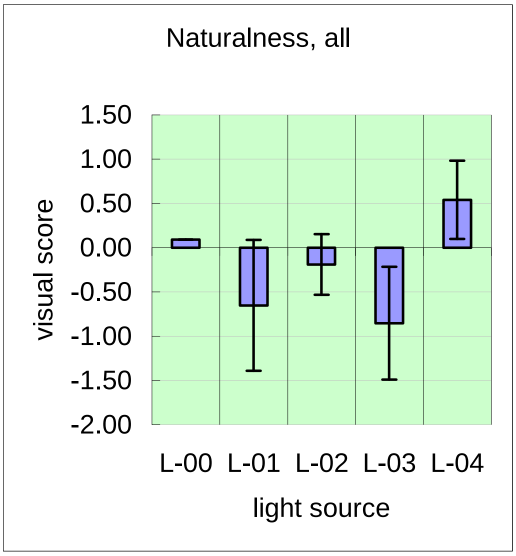


Figure 33. Pair-comparison, naturalness, all observers.

Also here we performed the rank order determination for the Inc., the CFL and the two p-LEDs.

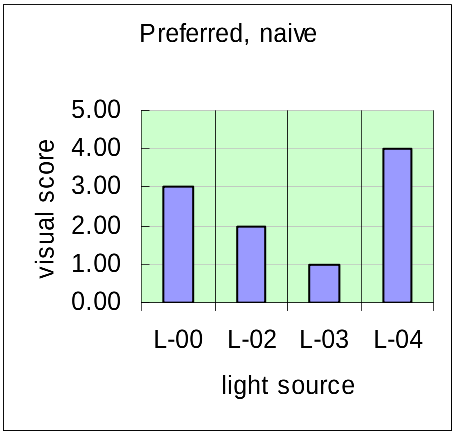


Figure 34. Rank order, pair comparison, preference, naïve observers.

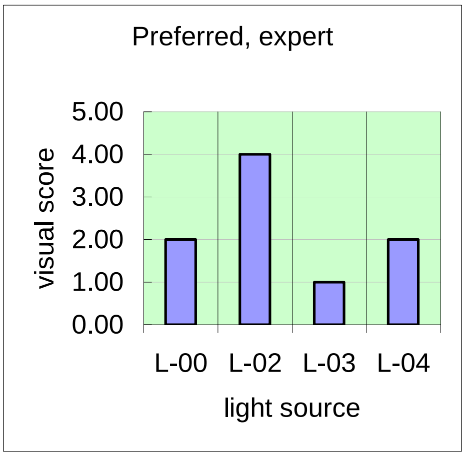


Figure 35. Rank order, pair comparison, preference, expert observers.

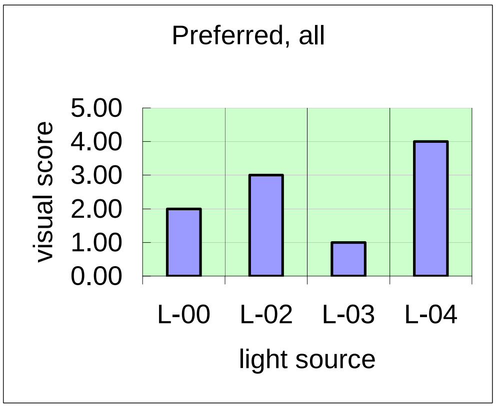
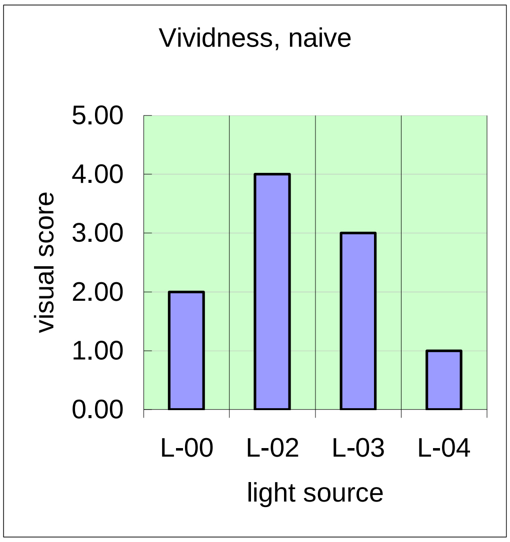


Figure 36. Rank order, pair comparison, preference, all observers.

One can see that for preferred the X-lamp scores high among the naïve and all groups.



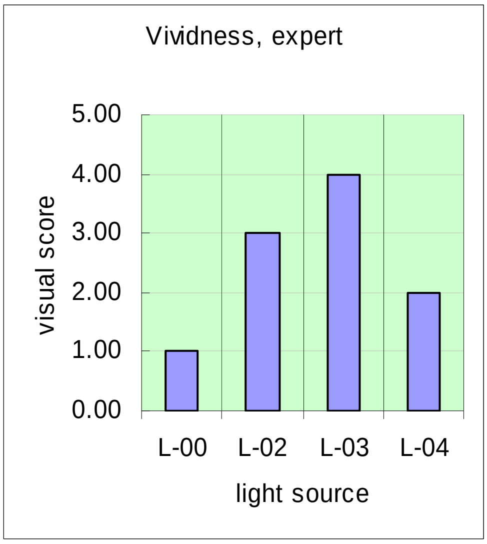


Figure 37. Rank order, pair-comparison, vividness, naïve observers.

Figure 38. Rank order, pair-comparison, vividness, expert observers.

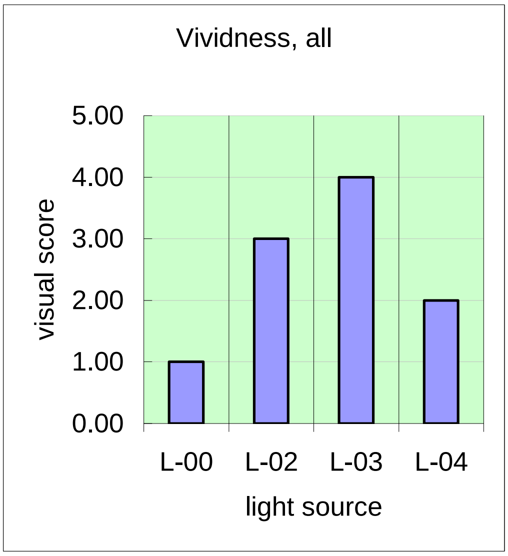


Figure 39. Rank order, pair-comparison, vividness, all observers.

Vividness is an exceptional descriptor, where observers do their decisions based on some unforeseen characteristics that should be further studied.

Naturalness is one of the most important characteristics for images, rank order results are seen on the following figures:

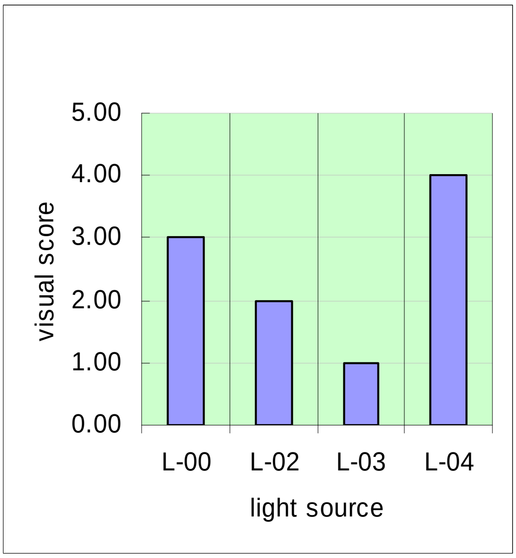


Figure 40. Rank order, pair comparison, naturalness, naïve observers.

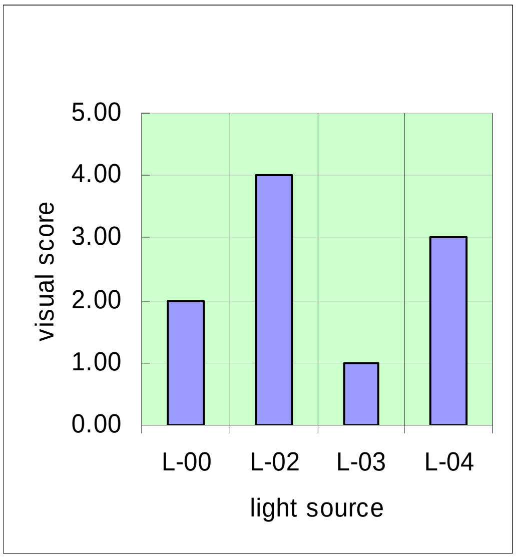


Figure 41. Rank order, pair comparison, naturalness, expert observers.

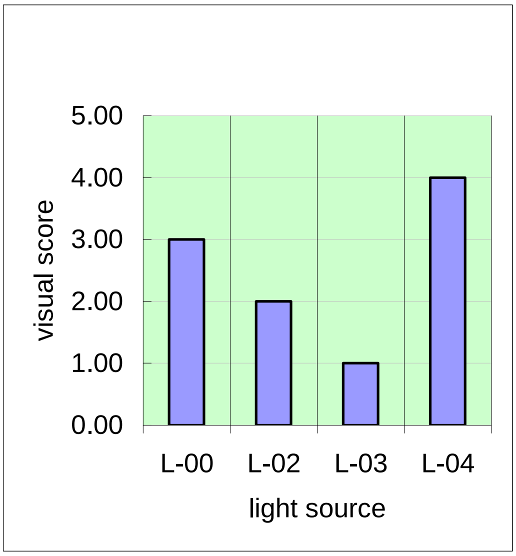


Figure 42. Rank order, pair comparison, naturalness, all observers.

The CFL preference for naturalness of the "expert" group is somewhat astonishing, as in everyday life they do not find the CFL a very natural light source. Both for the naïve and the total group the XE-lamp scores highest for naturalness.

The *colour fidelity* experiments were the most demanding for the observers. Here observers had to define the perceived colour difference between the colour of the ten Munsell samples seen in the test and the reference booths. The first eight Munsell samples, see Table 2, form a circle in colour space, thus for these samples one can expect a more or less continuous change of the observed colour differences. This is shown for the expert observers in Figure 43. Average visual colour difference values are depicted in Table 5, together with values of the naïve group, who's Munsell chip dependent data are seen in Figure 44.

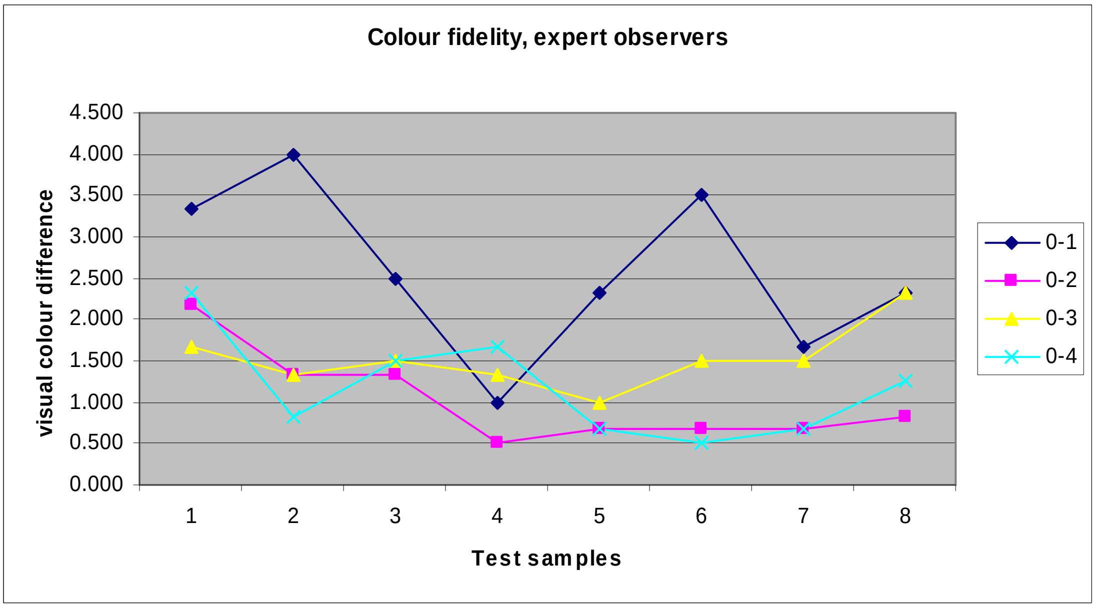


Figure 43. Visual colour difference for the first 8 Munsell samples comparing the observed colour under the test sources to the colour seen under incandescent light, expert observers.

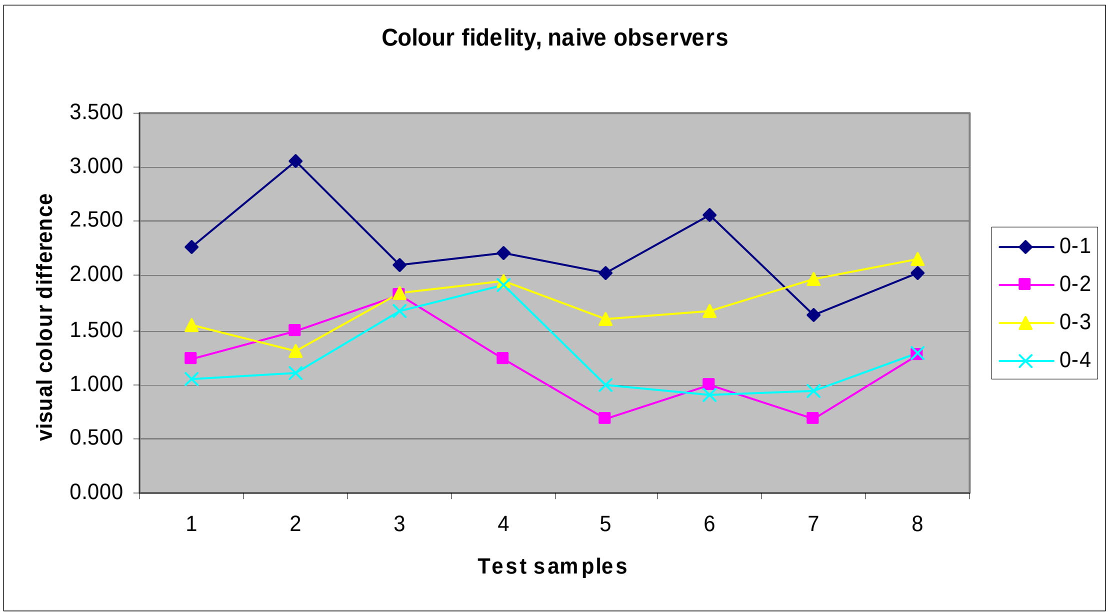


Figure 44. Visual colour difference for the first 8 Munsell samples comparing the observed colour under the test sources to the colour seen under incandescent light, naïve observers.

Table 5. Average visual colour differences for the eight Munsell test samples for the naïve and expert observers

Booth pair	Visual colour difference, naïve observers	Visual colour difference, expert observers
L-00 – L-01	2,23	2,58
L-00 – L-02	1,18	1,02
L-00 – L-03	1,76	1,52
L-00 – L-04	1,24	1,18

For both groups one can see that the CFL and the P/L LED-bulb provide both nearly equal colour fidelity, while the RGB-LED deviates most.

Summary

Colour fidelity and colour preference of four test light sources (a CFL, an RGB-LED and two phosphor-LEDs) have been compared with that of an incandescent lamp. Three types of visual tests have been performed: an absolute scaling of the colour of real fruits and vegetables seen under the five sources; a pair comparison of four ISO test images seen under the four test sources with similar images seen under the incandescent lamp light; and a colour fidelity test by determining the visual colour difference of ten Munsell chips illuminated by the incandescent reference source and the four test sources.

19 naïve and six expert observers, all of good colour vision, took part in the experiments.

Both direct evaluation of the single experimental set-ups and a rank-order determination which light source was preferred in the given situation has been prepared. Table **6** shows the result of the rank order evaluation.

Table 6. Overview of the rank-order results of the study

		Ranking			
		Incandescent L00	**CFL L02**	**Philips LED L03**	**XED lamp L04**
Single tests					
Vividness	all	2	1	3	4
	naive	4	1	2	3
	expert	1	2	4	3
Preference	all	3	1	2	4
	naive	4	1	2	3
	expert	1	2	4	3
Naturalness	all	3	1	2	4
	naive	4	2	1	3
	expert	1	2	3	4
Sum single all		***8***	***3***	***7***	***12***
Pair comparisons					
Vividness	all	1	3	4	2
	naive	2	4	3	1
	expert	1	3	4	2
Preference	all	2	3	1	4
	naive	3	2	1	4
	expert	3	4	1	3
Naturalness	all	3	2	1	4
	naive	3	2	1	4
	expert	2	4	1	3
Sum pair all		***6***	***8***	***6***	***10***

Ranking:
4 best
1 worst

(dark green) best rated in all three categories
(light green) best rated in a single category

Experiments have shown that experts and naïve observers evaluate colour distortions differently, but in all three experiments the P/L LED-bulb proved to be a good replacement of the traditional incandescent lamp, and worked better as the other-LED used as comparison.

References

1 CIE (1995) Method of measuring and specifying colour rendering of light sources New edition. CIE 13:1995.

2 CIE Technical Committee: Colour rendering of white light sources, TC 1-69.

3 CIE (1999) Colour rendering, closing remarks. CIE 135/2:1999.

4 ISO: Graphic technology – Prepress digital daa exchange – Part 2: XYZ/sRGB encoded standard colour image data (XYZ/SCID) ISO 12640-2:2004

5 Colorlite Kft Colour vision test