Title
ANALYSIS OF A VISUAL PERFORMANCE EXPERIMENT

Authors
Clear, R.
Berman, S.

Publication Date
1984-07-01

Lawrence Berkeley Laboratory
UNIVERSITY OF CALIFORNIA

APPLIED SCIENCE DIVISION

Presented at the Illuminating Engineering Society Annual Conference, St. Louis, MO, August 6-9, 1984

ANALYSIS OF A VISUAL PERFORMANCE EXPERIMENT

R. Clear and S. Berman

July 1984

APPLIED SCIENCE DIVISION

Prepared for the U.S. Department of Energy under Contract DE-AC03-76SF00098

DISCLAIMER

LBL-17763
EEB-L-84-12
L-94

To be presented at the Illuminating Engineering Society (IES) Annual Conference, St. Louis MO, August 6-9, 1984.

ANALYSIS OF A VISUAL PERFORMANCE EXPERIMENT

Robert Clear and Samuel Berman

Lighting Systems Research Group
Lawrence Berkeley Laboratory
University of California
Berkeley CA 94720 USA

July 1984

This work was supported by the Assistant Secretary for Conservation and Renewable Energy, Office of Building Energy Research and Development, Buildings Equipment Division of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

ABSTRACT

We reanalyze the Smith-Rea check value verification experiment. This experiment has been discussed in a number of articles, and is one of the 20 experiments used to support the CIE 19/2 model. A preliminary data sheet from Smith and Rea listed an incorrect score function and contained a large number of arithmetic errors in converting raw times to scores. Correction of these errors changes the CIE fit. We argue that the W_{123} parameter of this fit is not related to the "critical visual processes" as claimed.

We use the corrected data to examine basic trends. Subjects achieved their maximum scores for a large fraction of runs under all visibility conditions. There was no statistically significant difference in scores for tests from 100 to 5000 lux. Furthermore, illumination level was less important to performance than the other variables studied: subject, practice, and check set (legibility and contrast). The RQQ #6 recommended illumination levels for such tasks range from 200 to 750 lux, indicating that recommended levels may overstate the need for illumination.

There was a distinct practice effect, and this effect is correlated to visibility. The practice effect was largest where there was least visibility. The same set of checks was used in each run. It is not clear how much of the practice effect is due to this experimental artifact and how much can be generalized. The long-term magnitude of the visibility/performance trend is rendered extremely uncertain by uncertainty over the source of the practice effect. There is no question that there is at least a short-term visibility/performance trend.

The CIE regression is re-examined to see how efficient is its empirical description of the visibility/performance relationship. This analysis tests the hypothesis that even though the CIE model may not be theoretically correct, it may still be a good approximation. The four-point fit used in CIE 19/2 had only one degree of freedom and would be rejected if it was linear. Using less, or unaveraged data, we found that although the CIE fits explain a statistically significant amount of variance, they were less efficient than a simple ln(VL) fit.

It has been suggested that since VL is based on threshold contrasts, it is not an appropriate measure for supra-threshold real-world tasks. We performed a rank-order test of an alternative visibility measure, conspicuity, against performance, but found no correlation. As a hypothesis we suggest that visual performance is inherently bounded by threshold visibilities. There are several mechanisms that would lower nominally suprathreshold visibilities towards threshold levels in a visual performance experiment. The mechanisms are sufficiently different that there should be no unique visibility/performance relationship. Instead we argue that the relationship will depend on the type of the experiment (and hence the mechanism) and the details of the scoring function.

ANALYSIS OF A VISUAL PERFORMANCE EXPERIMENT

Paper 29.

Robert Clear and Sam Berman

Lighting Systems Research Group

Lawrence Berkeley Laboratory

University of California

Berkeley, CA 94720 USA

(415)-486-4094

I. INTRODUCTION

A previous paper criticized the formulation and use of the CIE 19/2 visual performance model, and voiced our concern about the statistical validity of the validation fits presented in the CIE report.[1-2] This paper re-analyzes the Smith-Rea Check-Reading experiment, uncovers numerical errors in previous reports on this experiment, and bears out our criticisms of the model.[3] We also analyze the data for basic trends, and examine it in terms of alternative models.

II THE CHECK-READING EXPERIMENT

The Smith-Rea check-reading experiment is data set 15 of the 20 data sets analyzed as validation for CIE 19/2.[1,4] A different interpretation of the results of this experiment is given in two papers by Ross.[5,6] A complete description of the experiment is contained in the original Smith-Rea report.[3]

The basic format of the experiment is that the subject compares 10 checks in succession against a printed list, which has from 0 to 3 errors. Forty conditions consisting of 4 illumination levels and 10 "readability" classes were examined. Four subjects made 4 runs for each condition. The subjects' pay was proportional to their score, S, where

$$S = H + f(T) - (M+FP) = 10 + f(T) - 2(M+FP) \quad (1)$$

and

$$f(T) = INT((35-T)/5) \quad T \geq 5 \quad (2)$$

$$= 0 \qquad T < 5$$

Where H is the number of correct comparisons, INT(x) takes the integer value of x, T is the time in seconds, M is the number of errors on the comparison list that the subject missed, and FP is the number of false positives (false identifications of an error). The time/score function, f(T), is set to zero for $T < 5$ to prevent the subject from simply flipping through the checks if they are hard to read. The data show that this was not a problem.

Preliminary hand-tabulated data sheets were sent to Ross for his analysis for the Federal Energy Administration. The data sheets listed a table incorrectly identified as total score that we call S',

$$S' = f(T) - (M + FP) \quad (3)$$

A comment in a follow-up letter that noted that the total score table was incorrectly identified was evidently missed. Ross analyzed the average value of S' per run as the total score. The CIE reports summed S' over subjects, check types, and runs to get 160 S' as the score at each illumination level. Only Smith and Rea reported the average value of S.

Although the difference between S and S' accounts for the major differences between the reports, we also found numerous errors in the calculated values of f(T) on the preliminary data sheet. Approximately 14% of the f(T) values were inconsistent with Eq. 2, significantly exaggerating the difference in performance between the highest and lowest illumination levels. The incorrect values were used in the Ross and CIE reports. They affected Smith and Rea's report only in their analysis of the frequency of maximum and minimum scores.

Ross used the data to show only that the bulk of the illuminance effect occurs at the lowest illumination levels. This conclusion is not seriously affected by the shift from S to S' or the errors in f(T). The CIE 19/2 model is supposed to provide a fit of "task performance, TP, selected by the investigator...".[1] Fitting S' instead of S exaggerates the effect of illuminance on performance by a factor of almost 4, and therefore seriously affects the regression parameters. Table 1 compares the CIE fit of S' to the equivalent fit to S. All three parameters of the fit are changed, and the changes to TP_{max} and W_{123} are statistically significant. We will comment on the indicators of goodness of fit (s^2 and R^2) later.

The Smith-Rea report provides a table of luminances and measured reference VL values for a sample of one check per readability class. An earlier version of this table in the appendix of Ross's FEA report has a relatively insignificant error in one of the luminance values, and nonsense VL values.[5] These VLs were not used in the Ross or CIE reports.

The VL values used in the CIE fit are calculated from the CIE formula, and, as shown in Table 2, differ from the measured or reference values. Measured and reference VLs are static values in that the target is viewed on axis (eccentricity $\tilde{X} = 0$), for a fixed time. The calculated values are supposed to apply to dynamic viewing conditions where neither viewing angle nor exposure time is controlled.

VLs depend on luminance, L; time; $\tilde{X}$; the critical size of the task, d, and the effective contrast, C. In visual performance experiments, $\tilde{X}$ is given as a function of the task demand parameter, D, and becomes an empirical correction factor for both time and eccentricity.

Effective contrast is the product of the equivalent contrast, $\tilde{C}$, and the contrast rendering factor, CRF. We calculated it from the measured (static) VLs.

The average over check and illuminance conditions, 0.774 ± 0.027, differs slightly from the CIE value of .764 (incorrectly identified as $\tilde{C}$). The difference was insignificant to our analysis.

The CIE 19/2 calculation used d = 4 minutes of arc in this and all other situations where d was not, or could not be, measured. This assumption is acceptable here, as the calculated Cs with d = 4 show no trend with illuminance.

VL values are moderately sensitive to d. The ratio of maximum to minimum VLs in Table 2 vary from 2.7 to 4.6 if d is varied from 1 to 10 minutes of arc. However, the actual fit is relatively insensitive to d. The variance, s^2, ranges from 11.8 to 14.1, and the fitted parameters vary by less than a standard deviation. This is due to the flexibility of the CIE formula and the saturation of performance at high VL.

Although the CIE 19/2 procedure uses dynamic VLs, static VLs were used to fit data sets 8 through 11. Furthermore, the Smith-Rea fit does not show that dynamic VLs are superior to static VLs. We suggest that the utility and correctness of the static/dynamic distinction should be examined more closely.

The number and significance of errors we found in just one experiment indicate that there may be substantial errors in the other fits as well. Equally significant is that the ability of the CIE model to fit incorrect data is consistent with our contention that it represents curve-fitting and not model-fitting. Note that the parameters D and W_{123} supposedly have physical interpretations in terms of visual processes.[1] Since speed and accuracy in this experiment were close to their maximum values at all visibility levels, fairly drastic changes in the score function makes little difference to the underlying physical measures of performance. On the other hand, Table 1 shows clearly that a change in score function can dramatically change the fitted parameters. This contradicts the direct visual interpretation of these parameters and indicates that the score function must be considered in the fits.[2]

The following section examines the data more thoroughly before returning to the problem of modeling the visibility/performance relationship.

III. BASIC TRENDS

The primary trends in the data are:

1) Subjects achieved their maximum accuracies and scores on a large fraction of runs under all visibility conditions.

2) There was a practice effect that is relatively large and inversely correlated to the visibility of the task. The practice effect may be an artifact of the experimental procedure.

3) There was no statistically significant difference in performance from 100 to 5000 lux.

4) Check type, subject, and practice level were substantially more important to performance than was light level. Legibility appeared to be more important than contrast and size. Finally, the correlation of VL to performance is larger than that of light level to performance.

Smith and Rea's results are similar but not identical to ours.[3] Their analyses of variance tested two classifications (two-way AOV): illumination level and check type, against four measures of performance: time, T; hit rate, H; false positive rate, FP; and total score, S. The only test that was not statistically significant was the number of hits as a function of illumination level. A practice effect was noted in the mean scores, but was believed to be unimportant. A rank-order test of performance was significant against readability, but not VL. Smith and Rea's frequency of maximum scores (=15) as a function of illumination level are incorrect, and were evidentally based on their preliminary tables of time points (discussed in Section II). Their values, 11, 33, 39, and 43, seem to have a positive trend over the entire illuminance range. The corrected values 19, 34, 34, and 37 are noticeably different only at the lowest illumination level (10 lux), which is consistent with mean value comparisons. There were also some minor errors in their other values and frequencies.

We ran four-way AOVs including run number and subject. The residuals were not normally distributed, and their variance was slightly larger than average for check set 1 and at low scores. Two residuals were anomalously large (>5 standard deviations from the mean). The time and total score tests, plus the subject and check set classifications tests against accuracy (H-(M+FP)), were better than 1% and are therefore probably significant despite the problems with residuals. Run number against accuracy was borderline significant at an approximate level of 5%. Illumination level was not significant against accuracy.

The illumination level/accuracy test should be considered inconclusive rather than negative. The data set is limited (124 errors total) and not normally distributed. The effect may be small since speed can be traded for accuracy. Finally, it should be remembered that Smith and Rea found a correlation against false positives.

Table 3 shows the results of using the "Duncan New Multiple Ranges" procedure to show which means in the AOV were significantly different for times and total scores.[7] The levels are ordered with performance increasing to the right. A line connecting values indicates that there is a greater than 5% probability that the mean values are the same. Overlapping lines mean that the test cannot distinguish between adjacent values--but that the end-points are significantly (less than 5% probability) different. This situation arises because there is insufficient data to show where the true breaks are. The major trends are examined in more detail below.

1) Maximum scores

Subjects reached the maximum possible score almost 20% of the time. However, in terms of the subjects's own maxima, almost 60% of the total and time scores were maxima. In addition, more than 80% of the runs were 100% accurate. Even at the lowest illumination level, 40% of the scores were at the subject's maximum. The 100% accuracy level is a physical limit, hence these high levels of performance show the relative ease of the task, and are not simply artifacts of the scoring function. This point is significant to applications professionals in that the experiment is representative of a moderately difficult office task.

2) The practice effect

Run 4 minus run 1 is a measure of the magnitude of the practice effect. A three-way AOV against this data provides a simple, although inefficient, test of whether practice affects relative results. At the 5% significance level, illumination and subject were significant against time and time points, and check set was significant against accuracy and total score, and just missed significance against time.

We ran rank-order regressions of the total score, time, and accuracy practice effects for the 40 check-illumination level conditions against VL to test for a correlation against visibility. The correlation coefficients were 0.381, 0.352, and 0.263, respectively. They are significant at the 1%, 2%, and 5% level (one-sided test). Since the correlation is significant, its magnitude is important. The relative scores against illumination level for runs 1 and 4 show that the effect is large: 0.957, .0989, 1.005, and 1, versus 0.981, 0.995, 1, and 1. The average over all runs (see Table 1, Section II) is closest to run 1. Extrapolation of the above trend as a linear function of RN/(1+RN), where RN is the run number, indicates that there may be no significant differences in performance once the subject has had sufficient practice.

The interpretation of the practice effect is complicated by the fact that the subjects examined the same set of 400 checks on each run. Use of the average over runs is equivalent to assuming that this repetition of checks is the dominant factor in the practice effect. Thus the long-term visibility effect may be less than the average values indicate.

3) Illumination level

Only the lowest illumination level, about 10 lux, gives significantly different, and lower, performance. The results are sensitive to experimental error. One-fourth of the performance improvement from 10 lux to 100 lux is due to just two outliers out of 160 points.

The performance trend is consistent with the CIE regressions. However, it is too weak to confirm the view that performance is monotonically related to illuminance. The data also do not provide evidence for a monotonic trend against VL. A three-way AOV of illumination level 4 minus illumination level 1 showed no sign of a check set (VL) interaction effect. We believe that the results add force to Ross's questioning of the performance gain from 100 to 5000 lux in that this gain may not be real.[6]

4) Check type

Performance with check sets 1, 2, and 5 are noticeably lower than with the others. Check set 1 contains checks with poor handwriting while check set 2 had low contrast or small sizes. Performance on check set 1 was significantly lower than on all other sets. Half the difference between check set 2 and the others was due to the two outliers mentioned earlier. Check set 5 had an intermediate readability rating, and the highest visibility rating. However, one check in this set was ambiguous and was responsible for the entire difference between this check set and the remaining sets. This analysis shows that legibility, as determined by penmanship, may be more important to visual performance than visibility, as determined by contrast and size.

This dominance of legibility over visibility is probably a major factor in why readability, but not VL, was statistically significant in Smith and Rea's rank-order correlation test. This was not a strong test, as there were only 10 categories. Note also that VL was determined for only 1 out of 40 checks. In a rank-order correlation test against the 40 check/illumination level scores, VL gives a better fit than illuminance alone ($r = 0.473$, $P \approx 0.2\%$, versus $r = 0.329$, $P \approx 5\%$). This VL/performance trend is signficant, but since there is a VL/luminance trend it does not prove causality.

IV. THE CIE 19/2 REGRESSION

The CIE model does not include all relevant variables, so it should be used cautiously even as an empirical model. All the variables, not just VL, must be close to those in the fit before the model will reliably predict performance.

The 20 CIE fits appear good, but data averaging makes them look better than they are. The CIE 19/2 model fits only relative performance, but in the Smith-Rea experiment, for example, performance varies with subject, practice level, and check type. Averaging over these variables provides a method of fitting the model to the data despite this problem.

Averaging loses information about the shape of the function and the power of the fit. In our example, averaging over check type eliminates all the information about how well the model handles the 10 contrast levels. In fact, averaging over subject, practice level, and check set reduces 160 data points to a single average value. The resultant 4-point, 3-parameter fit has only one degree of freedom. It is impossible to judge such a fit by eye. A multilinear fit of this type must "explain" (R^2=) 99.75% of the variance of the data to be considered statistically significant.

The CIE 19/2 model is nonlinear, so the above analysis does not apply directly. It is discouraging, however, to find that a simple log-linear fit, a + bln(VL), explains almost as much of the variance as the vastly more complicated CIE fit (R^2 = 0.995, and 0.996, respectively). The ln(VL) fit has two degrees of freedom and is statistically significant ($P \approx 0.5\%$). In fact it is more efficient than the CIE fit in that the estimated standard deviation from the ln(VL) fit is 3, versus 3.75 for the CIE fit. This comparison is not completely fair to the CIE model because there are so few degrees of freedom, but this latter problem is endemic to fits in the CIE report.[1,2]

We can increase the degrees of freedom of the fit by not averaging. Check sets differ in contrast. Since VL is proportional to contrast, the CIE fit should work with the data for individual check sets. However, check sets 1 and 5 had checks that were less legible than the other sets. These sets are excluded from the fit. This more than halves the variance of the fits, because the effect of legibility is not included in VL measurements.

Many of the CIE fits provide a separate value of the maximum score (TP_{max}) for each subject or group of subjects. In our example this gives six free parameters instead of three. Adding separate values of W_{123} and D for each subject is not effective and increases the variance of the fit.

We examined various runs and combinations of runs to isolate the practice effect (see Section III.2). The average over the runs is presented in Fig. 1. The scores on this plot were normalized by dividing by the appropriate TP_{max} so they could be shown against a single fitted curve.

The fit "explains" (R^2=) 42.6% of the original variance of the data, but about 3/4 of this explanatory power is due to the 4 TP_{max} values. The added explanatory power due to visibility is statistically significant, but the figure shows that there is little information on the shape of the visibility/performance relationship in this data. Even a ln(L) term adds statistically significant power to the fit, although the added explanatory power is only about 2/3 of the CIE visibility term. The term ln(VL) again leads to a more efficient fit than the CIE fit. The R^2 for the ln(VL) fit is 42.3%, but the variance is 0.1817 versus 0.1822 for the CIE fit. The complexity of the CIE fit does not improve its explanatory power.

The significance of the CIE fit is further reduced by the sensitivity of the parameters to practice level and to the two outliers. The value of W_{123} varies from 4.3% to 25%, while that of D varies from 52 to 88 for runs 3 and 4 versus run 1. Deweighing the two outliers reduces W_{123} by about a factor of two. The removal of check sets 1 and 5 from the analysis has little effect on the CIE fitted parameters, although it affects the shape of the performance versus illuminance trend, as follows: 0.968, 0.998, 0.993, and 1.0. The score for illumination level 2 is better than for illumination level 3, since check set 1/illumination level 2 had by far the lowest score of any check set/illumination level combination. The reason for the small effect on the fitted parameters is that half of the variation in ln(VL), and hence most of the information about shape, occurs at illumination level 1. The relative score at illumination level 1 is essentially unchanged by the deletion of data sets 1 and 5.

Even the significance of the correlation to VL is less than it looks. Again, most of this correlation is due to the correlation of VL with luminance. Excluding the points for illumination level 1 eliminates any ln(L) correlation and gives a ln(VL) term that is significant only at the 10% level. We believe that there are theoretical grounds for thinking that VL is related to performance; however, these data are at best a weak confirmation. The data do not show any significant difference between fits to static or dynamic VL, and there is not a sufficient range of VL in this experiment to clearly define the shape of the VL/performance relationship.

The ln(VL) fit cannot be correct at low or high VL, yet only 12 of the 20 CIE fits span a sufficient range to show this. Our experience with these data suggests that even these 12 sets can be fit with something simpler than the CIE function. Figure 1 shows that the fits are much less extraordinary than the CIE report makes them seem, and that the model implies far more than the data show.

V. OTHER MODELS

Studies of nerve-firing rates show that the retina responds roughly to the logarithm of contrast.[8] VL represents the ratio of the actual contrast of a task to its threshold (50% detection) contrast, so it can be considered a normalized contrast. A rough ln(VL) relationship to performance is therefore consistent with retinal studies.

Yonemura, however, suggests that apparent contrast (conspicuity) is a better measure of visibility than VL for real-world tasks.[9] VL requires extrapolating from threshold conditions, while conspicuity is measured at normal luminance and contrast levels. Conspicuity, however, is measured under steady viewing while visual performance experiments and VL involve a time constraint. Yonemura's data show that if the contrast is high enough, conspicuity drops when illuminance is increased. This trend is opposite to that of performance in the Smith-Rea experiment, and a rank-order test of estimated conspicuities against performance indicates that the two are not correlated ($r = -0.02$).

VLs in this and other visual performance experiments are sufficiently high (>2) that conspicuity should be applicable. CIE 19/2 implies that conspicuity is not applicable because eccentricity lowers effective visibility. This rationale is used to introduce the empirical shape-fitting parameter, D, into the CIE model.

Inditsky *et al.* proposed an explicit mechanism, based on the eccentricity concept, for the speed of detection of a target whose location is unknown. They note that the effective size of the area around the line of sight in which a target can be detected (the visibility "lobe") depends on VL and the threshold contrast as a function of eccentricity. They fit speed of detection to a probability model based on the area of the visibility lobe versus total area. This appears to be the first attempt to directly model speed.

However, it is unlikely that this mechanism is important in the Smith-Rea experiment. It should not take 1 to 3 seconds to find a target whose location is known. A more likely mechanism is the increased difficulty in identification versus recognition or detection.

Identification requires more information than does detection. Signal detection theory shows that this translates into a loss in effective visibility.[11] In the Smith-Rea experiment the numbers cover approximately 30' by 50'. If the critical detail size is approximately 4', there can be as many as 280 spatial channels of information. If each channel is independent, overall noise will go as $\sqrt{280}$.[11] Since the signal-to-noise ratio appears to be proportional to VL^2, the net reduction in effective VL would be 4.[10] Subjects can narrow their focus to a small area to improve effective VL and accuracy at the expense of the increased time needed to cover the target.

Since numbers have fairly well defined shapes, there should be fewer than 280 independent channels of information. Roughly 10 channels is consistent with the fact that only illumination level 1 shows a significant drop in performance. The mistakes at higher illumination levels are probably due to variations from the mean VL, legibility problems, or mental confusion. The last two mechanisms would provide a background level of mistakes independent of visibility. These speculations should be tested with more detailed findings.

The fact that Inditsky's proposed mechanism seems to work for one type of experiment, and a very different mechanism seens to work in this experiment indicates that there is probably no unique VL/performance relationship. The original CIE work involved measuring accuracy for a fixed exposure time. The generalization to performance was made without an explicit mechanism. The failure to explicitly consider speed, accuracy, the mechanisms driving them, and the trade-offs between them as determined by the score function, is what prevents the CIE model from predicting absolute, and to some extent relative, performance. A better understanding of VL/performance relationships requires explicit modeling of the mechanisms and conditions of the experiments.

VI. CONCLUSIONS

The CIE 19/2 data analyses do not confirm the functional validity of the model. Each experiment will have to be far more carefully analyzed to determine what information these data sets can contribute to our understanding of visual performance.

Our analysis of the Smith-Rea experiment shows that the CIE identification of VL as a fundamental visibility parameter is consistent with the data. The functional form of the CIE model is, however, not theoretically sound, and the actual fit is not even a particularly efficient empirical fit for this data. An attempt to model the problem more carefully indicates that there is no reason that a unique VL/performance function should exist. Instead it appears that there are a number of plausible relationships between VL and accuracy or speed, and that the actual VL/performance relationship will depend on the conditions of the experiment and the details of the score function.

On a more pragmatic level, our analysis of the Smith-Rea data shows that the visibility effect was small and indefinite. The young adults who were subjects in the Smith-Rea experiment showed very little (and possibly no) real improvement in performance above 100 lux. The RQQ #6 recommended illuminance level for young adults reading checks (levels D-E: 200-750 lux) shows a substantial safety factor built into the recommendations. The argument that this safety factor leads to a more flexible installation is no longer obviously valid. The increasing prevalence of video display units that require low ambient lighting for good visibility indicates that overlighting may produce a performance penalty. Experimental results should be carefully reviewed if future IES recommendations are to be more closely related to performance.

Acknowledgement

This work was supported by the Assistant Secretary for Conservation and Renewable Energy, Office of Building Energy Research and Development, Buildings Equipment Division of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

References

[1] Publication Commission Internationale de L'Eclairage, 19/2, An Analytic Model for Describing the Influence of Lighting Parameters upon Visual Performance, International Commission on Illumination, Paris, Frances, 1981.

[2] R. Clear and S. M. Berman, A New Look at Models of Visual Performance, Journal of the IES, 242-250, July 1983.

[3] S. Smith and M. Rea, Check Value Verification Under Different Levels of Illumination, Illuminating Engineering Research Institute report, 1982.

[4] H. R. Blackwell, Analyses of Twenty Sets of Visual Performance Data in Terms of a Quantitative Method for Prediction of Visual Performance Potential as a Function of Reference Luminance. Journal of the IES, 91-124, January 1983.

[5] D. K. Ross, Lighting and Thermal Operations--Energy Conservation Principles Applied to Office Lighting, Conservation Paper No. 18, Federal Energy Administration, Washington, D.C., April 1975.

[6] D. K. Ross, Task Lighting---Yet Another View. *LD&A*, 8(15), 37-43, 1978.

[7] H. L. Adler and E. B. Roessler, *Introduction to Probability and Statistics*, W.H. Freeman and Company, San Francisco, (1972).

[8] T. N. Cornsweet, *Visual Perception*, Academic Press, (1970).

[9] G. Yonemura et al., Equal Apparent Conspicuity Contours with Five-Bar Grating Stimuli, NBSIR 79-1925, National Bureau of Standards, U.S. Department of Energy, Washington, D.C., May 1980.

[10] B. Inditsky, H. W. Bodmann, and H. J. Fleck, Elements of Visual Performance, *LR&T*, 14(2), 218-231, 1982.

[11] J. A. Swets ed., *Signal Detection and Recognition by Human Observers*, John Wiley and Sons, Inc., New York (1964).

Table 1

Comparison: CIE 19/2 Fit of S´ and S

VL (CIE fit)	TP^a	S´ $\frac{TP_i}{TP_4}$	RTP	TP^a	S $\frac{TP_i}{TP_4}$	RTP
3.2	565	0.856	0.843	2147	0.965	0.953
6.8	633	0.959	0.945	2199	0.989	0.976
9.2	648	0.982	0.967	2213	0.995	0.982
10.4	660	1.000	0.985	2224	1.000	0.987

Parameters of fit

Data Type	TP_{max}	W_{123}	D	R^2	s^2
S´[b]	670	0.34	60	0.995	29
S´(L.S.)[b]	680±14	0.37±.05	66±8	0.996	21
S(L.S)	2250±25	0.115±.05	76±15	0.996	14

a TP = Task performance = Total points for the four subjects for 40 runs each.

b The CIE fits were done by eye. We redid the fits with a nonlinear least-squares (L.S.) program (S´(L.S.) above) to estimate the parametric uncertainities. To maintain consistency with the CIE fits we assumed that the variance of the data points were equal.

S´ Score as reported in CIE 19/2.

S Score as originally measured by Smith and Rea.

Table 2: Comparisons of VL Values

Luminance (cd/m^2)	2.44	31	251	1220
CIE VL values [a] (dynamic VL)	3.18	6.80	9.22	10.4
Reference VL [a] (static: $\bar{X}$ =0)	2.75	6.57	9.41	10.8
Measured VLs (static expt.)	2.74±.21	6.71±.48	9.41±67	11.3±.80
Reference VL [b] (static calculation)	2.79	6.66	9.53	11.0

a) Calculated at C = 0.764 (CIE value).

b) Calculated at C = 0.774 (best fit to Smith and Rea data).

Table 3

Comparisons of Mean Values: 5% Significance

Time Data

Classification	Ordered Rankings									
runs	1	2	3	4						
			(--	--)						
illumination	1	3	2	4						
		(--	--)							
			(--	--)						
check type	1	2	9	5	3	4	6	8	7	10
			(--	---	---	---	---	---	---	--)

Total Scores

Classification	Ordered Rankings									
runs	1	2	3	4						
		(--	--)							
			(--	--)						
illumination	1	2	3	4						
		(--	---	--)						
check type	1	2	5	3	9	6	7	10	4	8
		(--	--)	(--	---	---	---	---	---	--)
			(--	---	---	--)				

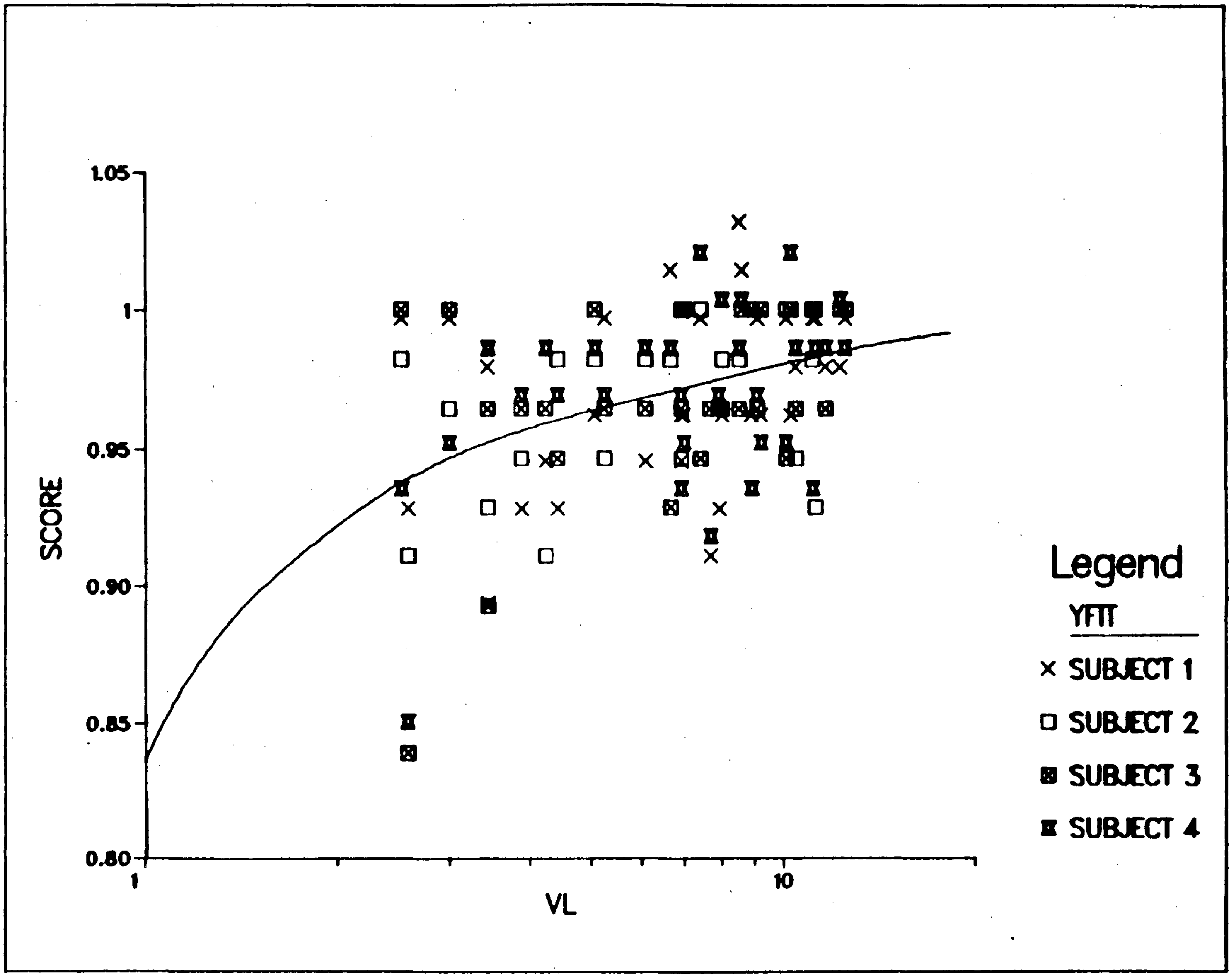

Figure 1. Normalized scores versus visibility level.

TECHNICAL INFORMATION DEPARTMENT
LAWRENCE BERKELEY LABORATORY
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720