Computationally rendered architectural spaces as means of lighting design evaluation

Hesham Eissa[1], Ardeshir Mahdavi[1], Roberta Klatzky[2], Jane Siegel[3]
[1]*School of Architecture, Carnegie Mellon University*
[2]*Department of Psychology, Carnegie Mellon University*
[3]*Human-Computer Interaction Institute, Carnegie Mellon University*

Key words: Lighting simulation, Computational visualization, Subjective lighting evaluation

Abstract: This paper describes an empirical study that was conducted to determine whether and to what extent subjective evaluation of lighting in architectural spaces can be reproduced using computationally rendered images of such spaces. The results imply that such images can reliably represent certain aspects of the lighting conditions in real spaces.

1. INTRODUCTION

Scientific simulation of light distribution in architectural spaces has been used in the past to generate numeric values of lighting performance indicators such as illuminance levels and daylight factors (e.g. on task surfaces), luminance levels, and glare indices. Designers and consultants typically compare such numeric results with minimum or maximum requirements in relevant illuminating engineering standards to decide if a particular design meets mandated performance criteria. Recently, it has been suggested that such traditional numeric evaluation methods may be complemented (or even substituted) by approaches that rely on scientific visualization tools, enabling the users to virtually observe the illuminated space and "directly" evaluate its lighting. The reasoning is that such scientific visualizations combine photo-realistic rendering with photometric accuracy, thus providing an image of the architectural space that is a dependable representation of its lighting. Consequently, designers could use computational visualization of lighting conditions in spaces in order to judge their visual quality. We consider this idea as a hypothesis in need of empirical testing. In this paper, we describe such a test.

2. PRIOR RESEARCH

Prior research investigated the effect of light on impressions of activity setting or mood. Also this research considered how we can recognize spatial lighting patterns as part of visual language that can assist the designer in implementing such impressions. Flynn tested the lighting-cue theory by accumulating data through *semantic differential scaling* and *multidimensional scaling* techniques (Flynn et al. 1973). The results showed how a system of visual cues could be recognized and interpreted in consistent ways by those who share a cultural background. This implied that it is possible to formulate standards for certain subjective aspects of lighting design.

In this context, visualizations that would resemble the real space could facilitate a more efficient and less costly procedure to investigate non-quantitative factors in lighting design. Hendrick et al. 1977 replicated Flynn's study to determine whether the subjective response to the visual experience of a space, as represented by two-dimensional slide projections, is comparable with the subjective response to real spaces. Their results suggest that there may be a potential to use (two-dimensional) slides of architectural spaces as representations of the lighting conditions in actual (three-dimensional) spaces. They also recommended that further research should establish a standardized method of photographing spaces to reproduce the effects of those spaces with little error (Hendrick et al. 1977).

In a more recent study by Veitch and Newsham (1996), the computational visualization tool Radiance was used to simulate different lighting settings in a workplace. The idea was to investigate the lighting designers' evaluations of lighting quality based on computer-generated renderings. The test lighting conditions were defined by a 3x3 matrix of levels of Lighting Power Density (LPD) and lighting quality as defined by designers (Designer' Lighting Quality, DLQ). The test of these lighting quality levels showed a remarkable disagreement in the overall ratings (overall lighting quality was rated on 1 to 9-scale), while semantic differential scale ratings showed better agreement. Veitch and Newsham recommended that future research should systematically investigate the *validity of rendering tools* both for lighting design and as means of representing lighting alternatives to clients. Also that study confirmed the reliability of *semantic differential scaling* as a subjective measure of a lit environment (Veitch and Newsham, 1996).

3. EXPERIMENTAL STUDY

3.1 Overview

The goal of the experimental analysis was to see if and to what extent the subjective lighting evaluation of computationally rendered images of spaces is consistent with subjective lighting evaluation of real spaces. The analysis involved the following steps:

a) Based on the results of previous research, a metric was selected that captures certain subjective light quality dimensions. This metric was then slightly modified based on the results of a bipolar semantic survey, and an experiment involving test participants evaluating lighting attributes of images of various architectural spaces.
b) Five actual lighting situations were selected involving three spaces (an office, a conference room, and a computer cluster) and different lighting schemes.
c) These situations were evaluated by a first group of test participants using the above-mentioned subjective lighting metric.
d) High-quality renderings of the above-mentioned situations were generated using an advanced visualization tool. The rendered images satisfied both the visual resemblance condition (sufficiently realistic depiction of the real lighting situations) and photometric reliability (sufficiently accurate replication of photometric measurements in the real spaces).
e) The rendered versions of the lighting situations were evaluated by a second group of test participants, using again the subjective lighting metric.
f) Subjective lighting assessments of the real spaces were compared with those of the computational visualizations to empirically determine the degree to which such visualizations can represent real spaces toward subjective lighting evaluation of architectural designs.

Sections 3.2 to 3.7 below describe these steps in more detail.

3.2 A subjective lighting metric

A common method to capture subjective impressions of otherwise non-measurable phenomena is the use of semantic differential rating scales. These include pairs of terms on a bipolar scale that addresses various dimensions of people's subjective impressions. For the purposes of this study, four steps were undertaken:

a) Suggestions for such bipolar pairs of terms were collected via a survey of 44 Architecture students;
b) The scales resulting from this survey were compared with the semantic differential scales developed by Flynn et al.1973;
c) Taking both the results of the survey and Flynn's scales, 28 pairs of terms were selected for further consideration;
d) This group of terms was tested using a small group of test participants who used it to evaluate the lighting quality of a number of office spaces as projected in slides. The final metric was derived based on a statistical analysis of the results of this test. Principal component analysis (PCA) was used as a data reduction method to eliminate the redundancy among the selected scales. The resulting scales are 10 pairs of terms under 7 categories (cp. Table 1). This differs only slightly from the scale adapted by Flynn et al. 1973. As an example, figure 1 illustrates one of the 7-point bipolar scales used in this study.

Table 1. The 10 selected semantic differential rating scales

Category	Terms
1. Evaluative (Psychological Impression)	Pleasant - Unpleasant
	Interesting - Boring
	Cheerful - Somber
	Shiny - Dull
2. Perceptual Clarity (Brightness)	Bright - Dim
3. Spaciousness	Large - Small
4. Light Distribution	Uniform - Non uniform
5. Spatial Complexity	Simple - Complex
6. Formality	Private - Public
7. Thermal, acoustic, and haptic associations	Cool – Warm

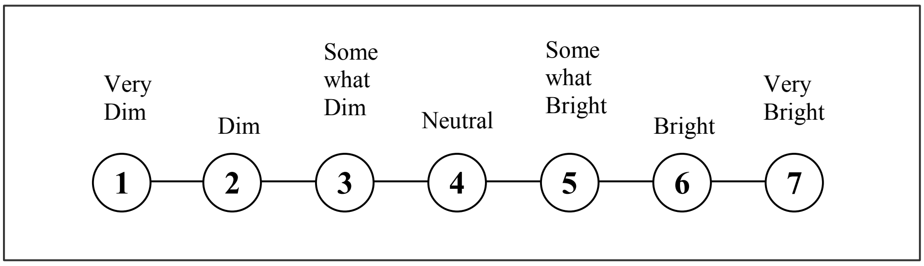

Figure 1. Example of a scale (Dim/Bright) of the derived metric (semantic differential)

3.3 Spaces and scenes

To capture people's subjective impressions of lighting, five actual lighting situations (scenes) were selected. These involved three spaces (an office, a conference room, and a computer cluster) and different lighting schemes (cp. Table 2). Schematic plans of spaces 1-A, 1-B, 2, 3-A, and 3-B are given in figures 2 to 4. Digital photographs of scenes 1-A, 2, and 3-A are given in figures 5 to 7.

Table 2. The five spaces and lighting schemes

Scene	Space	Lighting scheme
1-A	Office space 1	Indirect fluorescent
1-B	Office space 1	Spot – down light
2	Conference room	Indirect Halogen
3-A	Computer cluster (Large)	Direct fluorescent
3-B	Computer cluster (Small)	Direct fluorescent

3.4 Evaluating real spaces

A group of 50 people participated in the subjective evaluation of the above-mentioned scenes. The majority of the participants were senior architecture students, since this study was particularly motivated by the suggested potential of computational visualization as means of lighting design evaluation. The test participants observed the spaces from predefined vantage points (cp. the schematic plans, figure 2 to 4) and evaluated those using the semantic differential scale.

3.5 Renderings

The aforementioned five lighting situations were rendered using a commercially available visualization tool Lightscape release 3.2 (Lightscape 1999). The input information was derived either from direct measurements of the relevant attributes in the real spaces (e.g. reflectance of the walls) or extracted from documentation of the relevant equipment (light sources and luminaries). Figures 5 to 7 include the resulting renderings for scenes 1-A, 2, and 3A along with digital photographs of the corresponding (real) spaces.

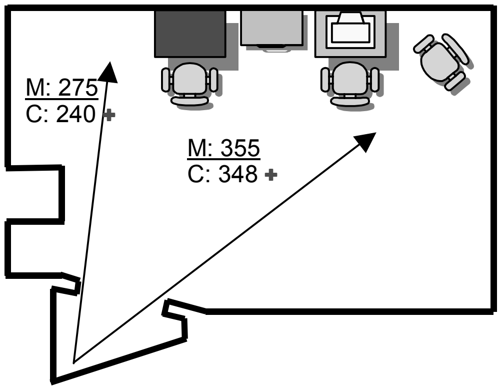

Figure 2. Schematic plan of space 1-A and 1-B (2 settings: fluorescent and down light) with indication of observers' vantage point as well as examples of measured (m) and computed (c) horizontal illuminance levels in lux (for the fluorescent lighting setting 1-A)

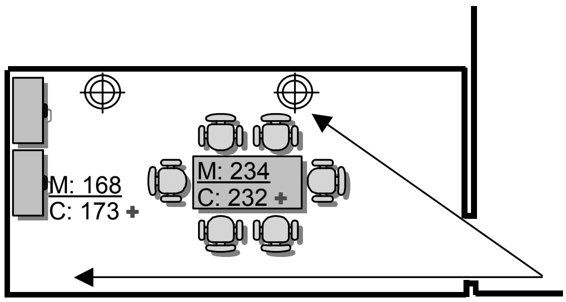

Figure 3. Schematic plan of space 2 with indication of observers' vantage point as well as examples of measured (m) and computed (c) horizontal illuminance levels in lux

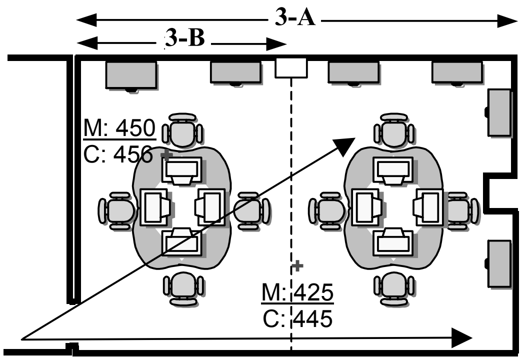

Figure 4. Schematic plan of space 3-A and 3-B with indication of observers' vantage point as well as examples of measured (m) and computed (c) horizontal illuminance levels in lux

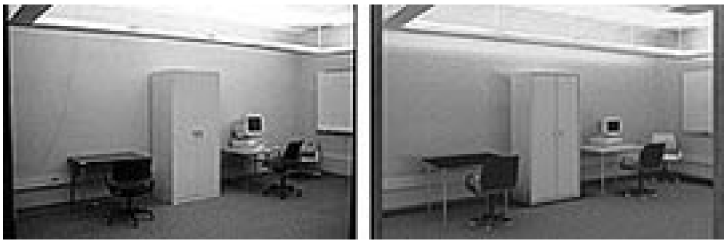

Figure 5. Space 1-A (Left: actual, Right: rendering)

Figure 6. Space 2 (Left: actual, Right: rendering)

Figure 7. Space 3-A (Left: actual, Right: rendering)

3.6 Evaluating rendered images

A second group of 50 people participated in the subjective evaluation of the five scenes based on their visualizations and used the same semantic differential scale. To avoid carry-over effects, no individual participated in both tests. Moreover, the profile of the second group of the test participants matched that of the first group in terms of age, gender, and educational background. The rendered scenes were shown to the test participants on a 19" colored computer monitor.

3.7 Results

The descriptive statistics for both tests are summarized in table 3. Participants' responses on a scale of 1 to 7 were recorded and the table shows the mean and standard deviation of their subjective impressions of actual spaces and the corresponding renderings. A subset of these results (mean evaluations for scenes 1-A, 1-B, and 3-A as compared to their respective renderings across the 10 scales of the semantic differential) is also illustrated graphically in figures 8 to 10.

Table 3. Descriptive statistics in terms of mean and standard deviation (in parenthesis) of evaluations of both actual scenes (S) and their renderings (R)

Scene:	1-A		1-B		2		3-A		3-B	
	S	R	S	R	S	R	S	R	S	R
1. Dim–Bright	4.6 (1.5)	4.9 (1.3)	2.1 (1)	1.8 (0.8)	4 (1.2)	3.1 (0.8)	5.7 (1.1)	5.6 (1.3)	6.2 (0.8)	5.8 (1)
2. Non uniform–Uniform	5.3 (1.5)	4.9 (1.6)	1.9 (0.9)	2.6 (1.3)	3.1 (1.6)	2.4 (1.3)	5.4 (1.4)	5.2 (1.6)	5.7 (1.3)	5.3 (1.3)
3. Boring–Interesting	3.0 (1.3)	3.4 (1.5)	4.8 (1.2)	4.4 (1.4)	5.3 (1.1)	4.8 (1.2)	4.1 (1.3)	3.5 (1.2)	3.6 (1.5)	3.0 (1.3)
4. Private–Public	5.0 (1.2)	4.6 (1.5)	2.6 (1.1)	2.9 (1.4)	3.7 (1.6)	3.5 (1.1)	6.0 (0.9)	5.7 (1.3)	5.6 (1.2)	5.0 (1.3)
5. Simple–Complex	2.7 (1.4)	3.2 (1.4)	3.4 (1.5)	3 (1.4)	4 (1.4)	3.4 (1.5)	3.6 (1.4)	3.4 (1.5)	3.2 (1.3)	2.9 (1.3)
6. Dull–Shiny	3.2 (1.2)	3.6 (1.2)	3.7 (1.2)	3.1 (1.3)	4.2 (1.2)	3.7 (1.4)	4.7 (1.2)	4.5 (1.4)	4.9 (1.4)	4.5 (1.5)
7. Small–Large	4.3 (1.1)	4.9 (1.3)	3 (1.2)	3.7 (1.4)	4.0 (1.3)	3.5 (1.1)	5.7 (0.7)	5 (1.4)	4.4 (1.3)	3.9 (1.3)
8. Unpleasant–Pleasant	3.7 (1.6)	4.2 (1.5)	4.8 (1.4)	4 (1.5)	5.3 (1.3)	4.8 (1.6)	4.6 (1.4)	4 (1.6)	3.9 (1.6)	3.4 (1.4)
9. Cool–Warm	3.6 (1.2)	3.8 (1.4)	4.7 (1.3)	4.4 (1.6)	5.1 (1.2)	4.6 (1.4)	3 (1.3)	2.7 (1.2)	2.9 (1.4)	2.8 (1.1)
10. Somber–Cheerful	3.4 (1.3)	4.2 (1.5)	3 (1.1)	2.8 (1.4)	4.2 (1.3)	3.9 (1.3)	4.4 (1.2)	3.9 (1.5)	4 (1.4)	3.5 (1.3)

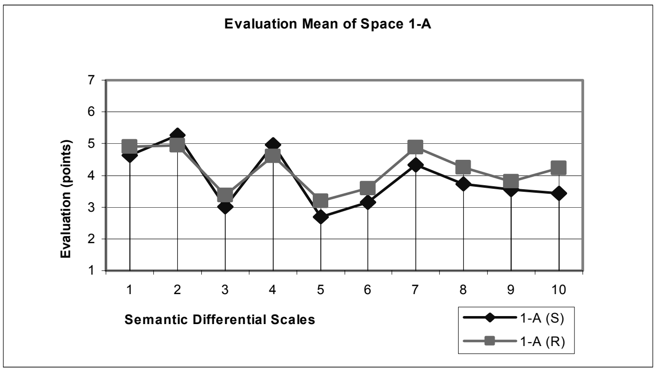

Figure 8. Evaluation means for scene 1-A
(S: actual scene, R: rendering, 1…10: scales according to table 3)

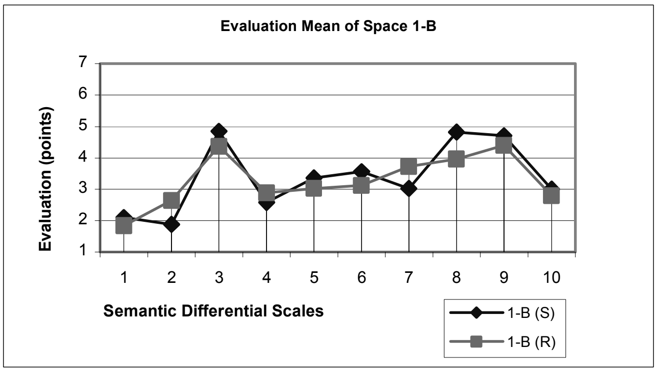

Figure 9. Evaluation means for scene 1-B
(S: actual scene, R: rendering, 1…10: scales according to table 3)

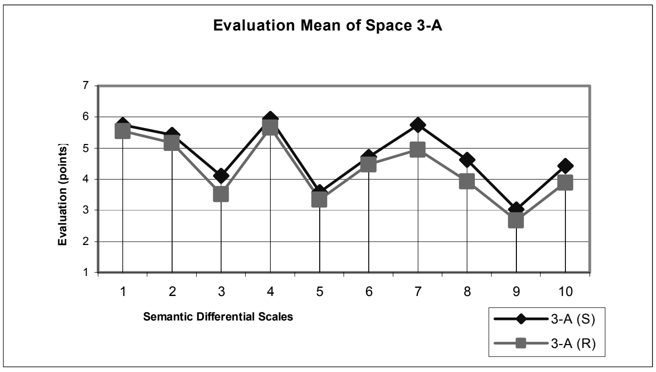

Figure 10. Evaluation means for scene 3-A
(S: actual scene, R: rendering, 1…10: scales according to table 3)

To establish an overall understanding of the degree of agreement between the results of the two tests, a regression analysis was performed (cp. figure 11) (McClave et al. 1997). The correlation (Multiple r) is found to be 0.91; the corresponding r2 of 0.83 indicates the variance accounted for by regression. Regression parameters are a slope of 0.81 (P-value 0.00) and intercept of 0.55 (P-value 0.02). These results imply a high level of congruence between impressions of the lighting gained from rendered images as compared to impressions gained from actual spaces, but with a relative reduction in the range of responses and constant offset of .55 for the rendered images.

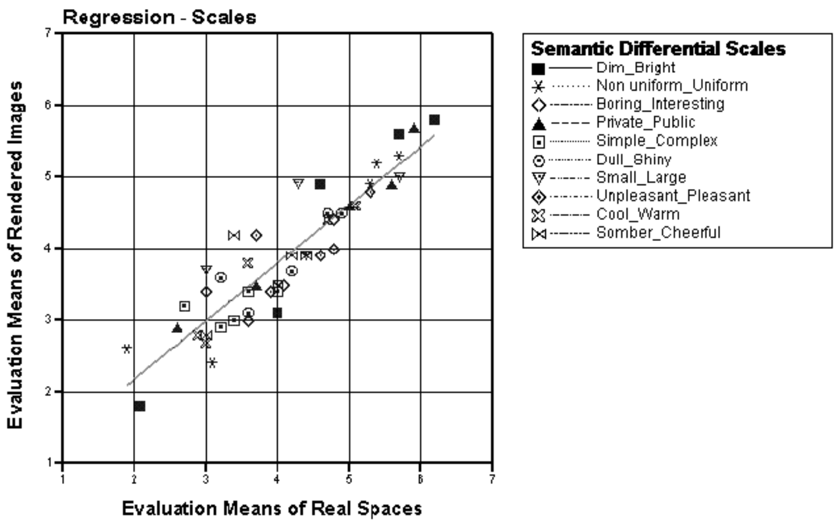

Figure 11. Regression analysis across all spaces and scales

Regression analysis was not only performed for the entire data, but also for individual scales. While the regression analysis of the entire data shows a high overall correlation, there are differences among individual scales in terms of their representational reliability (correlation) as well as in terms of their sensitivity to the stimuli variance (some scales display a wider range of numeric variability across different stimuli). The present study suggests that, while using rendered images to predict lighting quality features of real spaces, the dim/bright, non-uniform/uniform, boring/interesting, private/public, dull/shiny, and cool/warm scales are more reliable. Among this subset of scales, the dim/bright, non-uniform/uniform, and private/public display also a high sensitivity toward variance in stimuli.

In addition to regression analysis a non-parametric test (Kolmogorov-Smirnov) was used to compare the results of both evaluations. This test compares the evaluation medians for both data sets while considering each sample's distribution (Siegel and Castellan, 1988). The test was conducted for all 50 cases (10 scales and 5 lighting scenes). The resulting P-values of this test are summarized in table 4. Assuming a confidence level of 90% and a two-tailed test, P-values less than 0.05 would imply that there are significant differences between the evaluations of renderings versus actual spaces. As it can be seen from table 4, in 82% of the cases (41 out of 50), there is no significant difference between the subjective evaluations of actual scenes versus their rendered images.

Table 4. P-values of Kolmogorov-Smirnov test for all 50 cases, P-values less than 0.05 (in parenthesis) show significance at confidence level of 90% and two-tailed test.

Scene:	1-A	1-B	2	3-A	3-B
1. Dim–Bright	0.86	0.96	(0.00)	0.86	0.27
2. Non uniform–Uniform	0.39	(0.04)	0.11	1.00	0.54
3. Boring–Interesting	0.96	0.27	(0.04)	(0.04)	(0.01)
4. Private–Public	0.71	0.54	0.11	0.86	(0.02)
5. Simple–Complex	0.27	0.39	0.18	0.96	0.71
6. Dull–Shiny	0.27	0.54	0.54	0.71	0.71
7. Small–Large	(0.04)	0.07	0.39	0.07	0.18
8. Unpleasant–Pleasant	0.39	(0.01)	0.39	0.54	(0.04)
9. Cool–Warm	0.71	0.27	0.71	0.71	1.00
10. Somber–Cheerful	0.18	0.39	0.54	0.71	0.39

It is noteworthy that both the regression analysis and the Kolmogorov-Smirnov provide statistical evidence for the validity of the initial hypothesis.

4. CONCLUSION

The objective of the study described in the present paper was to empirically establish if and to what extent subjective lighting evaluation of lit architectural spaces can be reproduced using computationally rendered images of such spaces. For the sample of participants and lighting scenes tested in the study, the results suggest that such images can reliably represent certain aspects of the lighting conditions in real spaces. The overall correlation between the image-based and space-based evaluations was found to be significant. However, within the semantic differential metric, not all employed scales display both high representational consistency (from spaces to their rendered images) and high sensitivity (to variances in the stimuli, i.e. differences in lighting-relevant attributes of various spaces).

Future research should further investigate potential effects of *a)* alternative semantic differentials, *b)* multiple space use types, *c)* different samples of test participants (in terms of age, profession, and cultural background), and *d)* dynamic experience of real spaces as opposed to evaluations based on a single vantage-point.

Moreover, it became clear in the course of the present study that multiple conditions must be met, if rendered images are to be used as reliable surrogates of real spaces for lighting design evaluation. Such conditions include, for example, *a)* the application of proper scales for the lighting design task at hand, *b)* proper construction of the computational model (geometry, material properties, photometric data, etc.), *c)* application of a scientifically sound and empirically validated visualization tool, *d)* proper calibration of hardware (i.e. particularly computer display unit) using one or more real spaces as reference.

5. REFERENCES

Flynn, J. E., T. J. Spencer, O. Martyniuk, and C. Hendrick, 1973, “Interim study of procedure for investigating the effect of light on impression and behavior”, *Journal of Illuminating Engineering Society,* 3, p. 87-94.

Flynn, J. E., 1988, “Lighting-design decisions as interventions in human visual space”. In: J. L. Nasar (Ed.) *Environment Aesthetics: Theory, research, & applications,* Cambridge University Press, New York, p. 156-169

Flynn, J. E., J. A. Kremers, A. W. Segil, and G. R Steffy, 1992, *Architectural interior systems: Lighting, acoustics, air conditioning (3rd ed.),* Van Nostrand Reinhold, New York.

Hendrick, C., O. Martyniuk, T. J. Spencer, and J. E. Flynn, 1977, “Procedures for investigating the effect of light on impression: Simulation of a real space by slides.” *Environment and Behavior,* 9 (4), p. 491-510.

Lightscape, 1999, Autodesk release 3.2

McClave, J. T., F. H. Dietrich II and T. Sincich, 1997. *Statistics (7th ed.),* Prentice Hall, New Jersey.

Siegel, S., and N. J. Castellan, 1988, *Nonparametric statistics for the behavioral sciences (2nd ed.),* McGraw-Hill, New York.

Veitch, J. A. and G. R. Newsham, 1996, “Experts’ quantitative and qualitative assessments of lighting quality”, *The 1996 Annual Conference of the Illuminating Engineering Society of North America, Cleveland, OH, USA.*