# Nominal Scaling of Print Substrates

Nathan Moroney, Giordano Beretta

HP Laboratories
HPL-2009-322

**Keyword(s):**
digital printing, substrates, lightness, chroma, hue, thickness, roughness, fluorescence, opacity, glass, Nominal Scaling, perceptual categorization, corpus linguistics

**Abstract:**
Print substrates are often measured or characterized using optical measurements, such as spectra or colorimetric measurements. Specialized devices or techniques exist to measure other properties, such as thickness, surface roughness, fluorescence, opacity or gloss. However, these measurements are often disjoint and, with respect to color imaging requirements, they are often a secondary consideration. We present specific measurement data for a collection of digital commercial print substrates and explore the correlations of these measures and their general distributions. We then present the results of a nominal scaling experiment in which both the visual and tactile properties of 72 print substrates was evaluated by 21 subjects. These evaluations were based on unconstrained text descriptions of the samples. The analysis made use of techniques from corpus linguistics to determine multivariate clustering by keywords and by samples. This allows both a global view of the visual and tactile categorization schemes of the subjects, as well as specific pairing of print substrates deemed to be most similar by subjects. We conclude with a discussion of nominal scaling and it's relevance to perceptual categorization.

External Posting Date: September 21, 2009 [Fulltext] Approved for External Publication
Internal Posting Date: September 21, 2009 [Fulltext]

To be presented at the 17th IS&T SID Color Imaging Conference, Albuquerque, NM. Thursday November 12, 2009.

# Nominal Scaling of Print Substrates

***Nathan Moroney and Giordano Beretta***

***Hewlett-Packard Laboratories; Palo Alto, CA, USA***

## Abstract

*Print substrates are often measured or characterized using optical measurements, such as spectra or colorimetric measurements. Specialized devices or techniques exist to measure other properties, such as thickness, surface roughness, fluorescence, opacity or gloss. However, these measurements are often disjoint and, with respect to color imaging requirements, they are often a secondary consideration. We present specific measurement data for a collection of digital commercial print substrates and explore the correlations of these measures and their general distributions. We then present the results of a nominal scaling experiment in which both the visual and tactile properties of 72 print substrates was evaluated by 21 subjects. These evaluations were based on unconstrained text descriptions of the samples. The analysis made use of techniques from corpus linguistics to determine multivariate clustering by keywords and by samples. This allows both a global view of the visual and tactile categorization schemes of the subjects, as well as specific pairing of print substrates deemed to be most similar by subjects. We conclude with a discussion of nominal scaling and it's relevance to perceptual categorization.*

## Introduction

We start by considering the task of specifying media for commercial print jobs. This can be solved creating a catalog in which the available media options are listed, along with the color, geometric and mechanical appearance properties. We study how such a catalog or chart can be created. Because each print shop keeps a different media stock, we require a tool or methodology that can easily generate a catalog from any collection media stock. Current measurement devices and techniques for quantifying perceptual properties of print substrates tend to be univariate and specialized. Superficially, it might seem that there is only minimal perceptual variation in print substrates. An alternative possibility is that an integrated multimodal perceptual scaling of print substrates is simply too complex for conventional psychometric threshold and scaling techniques. We therefore seek to derive a global, yet specific, clustering of print substrates which considers all of the visual and tactile properties of print substrates.

The role of colorimetry in color management and color reproduction is well established and there is a growing understanding of how spectral data could be used for multispectral reproduction.[1] The perception and scaling of gloss has seen recent research results but these have been based on computer simulations[2] and a single black gloss ramp.[3] Gloss has also shown to be significant for object recognition[4]; commercial gloss meters and standards[5-7] exist for multi-angle gloss measurements. Fluorescence has seen recent research results[8,9] related to modeling and measurement. However, bi-spectral measurements are relatively rare. Opacity[10,11] can be measured using a color difference relative to a white and black backing. Thickness is a basic property that can be measured directly, although this is not necessarily the same as basis weight, which is related.[12] Finally, surface roughness[13] can be measured with a stylus profilometer[14] and while there are numerous possible derived metrics, the standard deviation relative to the mean or $R_a$ is used below.

To demonstrate roughness and gloss properties for the 72 commercial print media in our print shop's stock, the surface roughness was measured with a TIME TR200 Roughness Tester. In addition, the gloss units at 20°, 60° and 85° were measured with a BYK-Gardner micro-TRI-gloss meter. To visualize the range and variation in these values, Fig. 1 shows a scatterplot[15] of the measurements. There is some correlation for the lower gloss readings but in all cases the correlation breaks down for higher values. There is also some tendency for very smooth surfaces to have large gloss readings but there is some variation and a non-linear transform may be applicable. Note that the gloss readings for the transparent substrates have been clipped to 100 gloss units to allow for more spread to be visualized in the scatterplot.

A scatterplot of thickness, fluorescence, lightness and opacity is shown in Fig. 2. The thickness is in mm and was measured with Mitutoyo Digimatic digital calipers. The fluorescence was measured using a Minolta Chroma Meter CS-100 and based on the simplified technique in ref. 16. Essentially, this data is the luminance in cd/m$^2$ of the samples, given only UV illumination. The lightness is CIELAB $L^*$ and was measured with a GretagMacBeth Eye-One. The opacity is $\Delta E^*_{ab}$ between measurements with a CERAM Research Glossy White tile backing versus a CERAM Research Glossy Black tile backing. This scatterplot shows data with considerable spread, with significant concentration of values and extreme values. In qualitative terms, note that low fluorescence doesn't necessarily mean lower lightness and thinner media does not necessarily mean greater opacity. Note that the opacity values for the transparent media was clipped to 10 $\Delta E^*_{ab}$ to aid with the visualization of opacity.

The CIELAB chroma and hue values were also measured with the GretagMacbeth Eye-One. The result was ten dimensions of data, of which only lightness, chroma and hue have well defined psychometric scales. The scatterplots in Figs. 1 and 2 show a large amount of variation in the other 7 dimensions. Additional analysis was performed on this measurement set, but clearly the perceptual significance of all this data is ambiguous. How do human subjects actually categorize these properties? Can we better understand the visual and tactile perceptual correlates of these print substrates in an integrated manner? If we can answer these questions, we can produce more customer-friendlier stock catalogs. In addition given a sufficiently detailed sampling of these categories and samples measurements, what correlations can be made between specific categories and measurements?

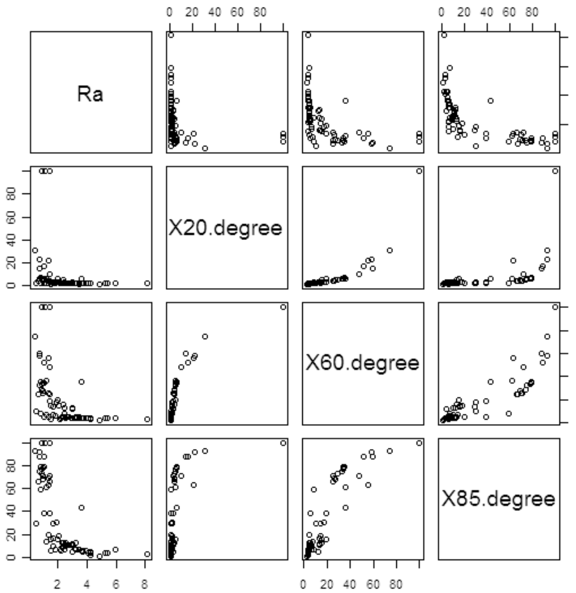


*Fig. 1. Scatterplot of surface roughness, Ra in microns, and 20°, 6°0 and 8°5 gloss measurements in gloss units.*

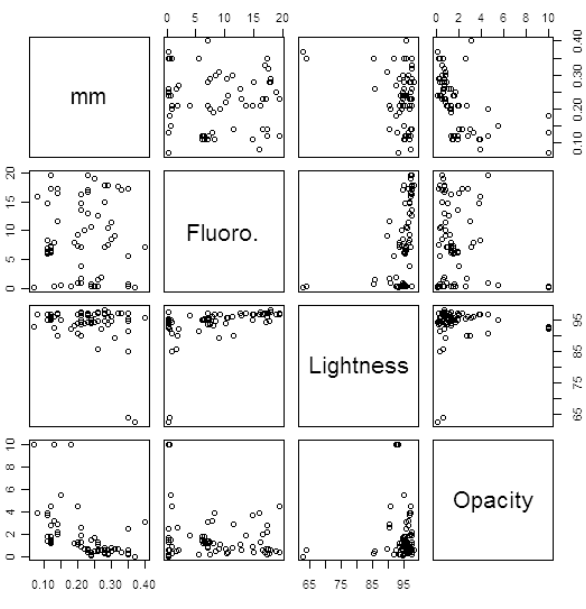


*Fig. 2. Scatterplot of thickness in mm, fluorescence in cd/m$^2$, CIE L* and opacity in $\Delta E^*_{ab}$.*

The results shown in Figs. 1 and 2 demonstrate that there is no lack of physical measurements to be made for print substrates. As an alternative or supplement to this data we will consider perceptual categorization of the substrates through unconstrained anchored elicitation. The term unconstrained means that no fixed vocabulary was defined in advance of the experiment. The term anchored elicitation means that words or descriptions were collected on a sample-by-sample basis give direct physical evaluation of the sample. The result is a collection of nominal assignments. This collection could also be considered a corpus. Corpus linguistics[17-19] is a methodological approach to empirical linguistics. A corpus is a machine-encoded or digitized representative sampling of text or speech. Representative sampling is a key consideration, although it is often considered with respect to the question under consideration or the research being conducted. In our case we are creating essentially a specialized corpus so we hypothesize that the modest size will not be a significant limitation.

Corpus linguists have derived a series of analysis tools and procedures that include pre-processing of the corpus, for example stemming or identification the lexeme or root word that would be looked up in a dictionary. A range of multivariate analysis and visualization techniques have also been applied, such as dendrograms. These figures represent clustering of samples or words based on increasing similarity. A dendrogram is essentially a tree where the leaves are most similar and the higher branches less similar. More specifically a dendrogram is a method for creating a two-dimensional visualization of pair-wise recursive distances within a set of higher dimensional data points. There are various hierarchical clustering algorithms but for this paper we will consider bottom-up or agglomerative clustering.

## Experiment

Using the 72 commercial print substrates measured in the introduction we created a visual and tactile test target. These print substrates consisted of a wide range of media including: plain papers, glossy papers, transparent media, metallic papers, mailing labels, magnetic sheets and other samples. The samples were anchored in a fixed sequence based on an alphabetic ordering by vendor. A per-subject randomization would have been preferable but given the large number of samples, observers and the desire to have all samples available for simultaneous evaluation, this was deemed not feasible. These targets are shown in Fig. 3; the subjects evaluated the samples in a GTI ColorMatcher Color Matching Booth with D50 illumination. A total of 21 subjects participated and consisted of a mix of print professionals, engineers, academics, and naïve participants. The specific instructions given to the subjects were as follows:

"Please evaluate the following 72 media samples. Specifically provide the words or phrases that best describe the visual and tactile properties of these samples. The larger sample can be touched for the tactile evaluation of the samples. The smaller target may be useful for evaluation of the visual properties but please do not touch it. For context imagine you are describing differences between the samples over the phone to a work colleague. Please provide at least one word per patch and separate your words or phrase with commas. You may work in any order, can use relative words or phrases, can use the same word or phrase multiple times and can revise your responses at any time. Thank you for your time — any questions?"

*Fig. 3. Visual target, left, and tactile target, right, in the viewing booth. The visual samples were 15 mm square and the tactile samples were 35 × 20 mm and were attached by their left edge to the foam core base.*

## Results

The observer data was recorded in a 9 column by 8 row table on an 11 × 17 inch piece of paper. This data was then transferred to a computer for analysis. Subjects varied in the amount of time required to provide unconstrained text descriptions of the samples, but roughly averaged 45 minutes. Given the machine-encoded text descriptions, for the analysis we followed the basic methodology of corpus linguistics. In this case all of the subjects' responses were pooled into a single corpus of roughly 7000 words with over 450 tokens. This text was then cleaned by conversion to all lower case and by spell checking. Minimalist stemming was applied by merging common terms ending in –s and –y. A sorted list of frequency of occurrence by proportionality was then computed. Words like 'white' and 'smooth' were perhaps not surprisingly used by many subjects for many samples. More surprisingly is the large number of additional words used. There were more than 20 words with a relative frequency of greater than 1%. There were also over 150 hapax legomena or words used only once. Clearly, while there is a superficial abundance of smooth white print substrates, there is also a lexical richness that is both specific and diverse.

The initial analysis reveals that, subjects are not simply using a small number of words to describe print substrates. The next step is to explore how these terms are interrelated. Are there terms that are roughly synonymous? Given the design of the experiment, the corpus is highly structured and is essentially an anchored elicitation task. It is possible then to create a matrix of occurrence based on keywords and samples. Specifically if we consider the top 64 keywords and the 72 print substrates, it is possible to create a 64 by 72 matrix where each entry in the matrix is a count that corresponds to the number of times that keyword was used with that substrate sample. This relative frequency can then be converted to a correlation matrix, in this case using the Kendall tau rank correlation coefficient. Finally, it is possible to compute a distance matrix and apply agglomerative clustering to cluster similar keywords. This clustering can then be visualized as a dendrogram, shown in Fig. 4.

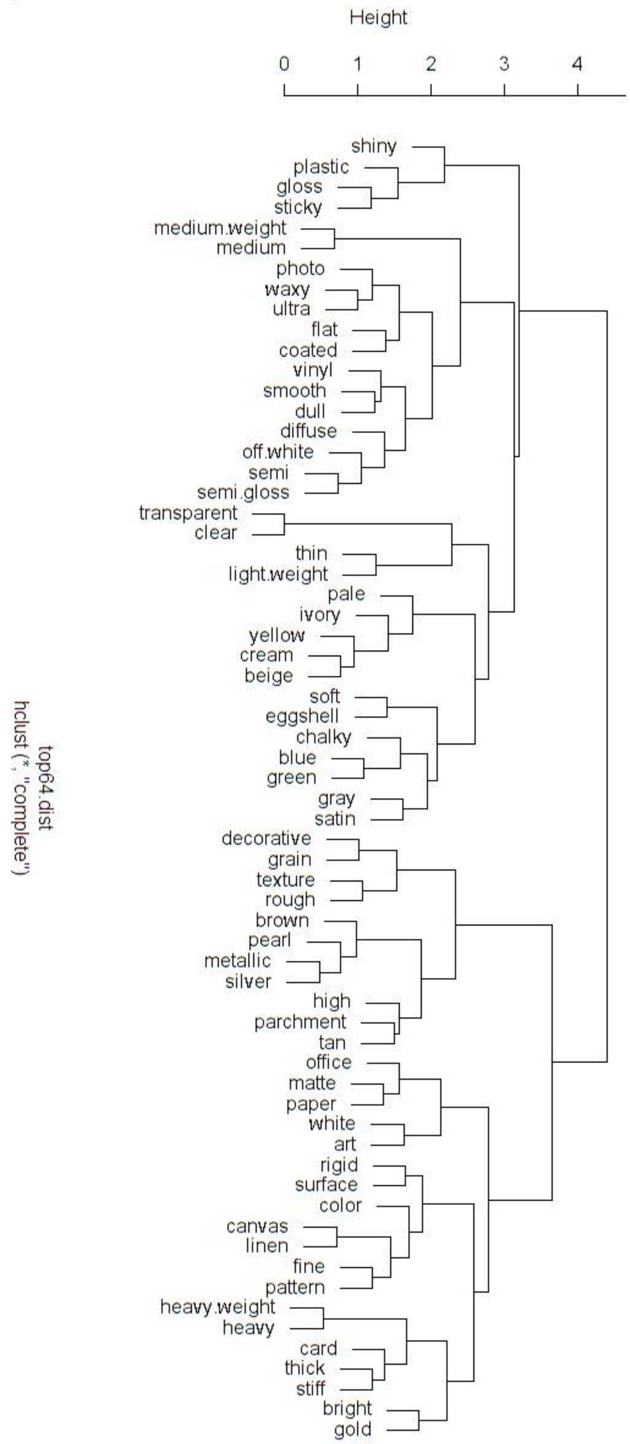


*Fig. 4. Agglomerative clustering of the top 64 keywords based on incidence of usage across the samples using Kendall's tau rank correlation coefficient and visualized as a dendrogram.*

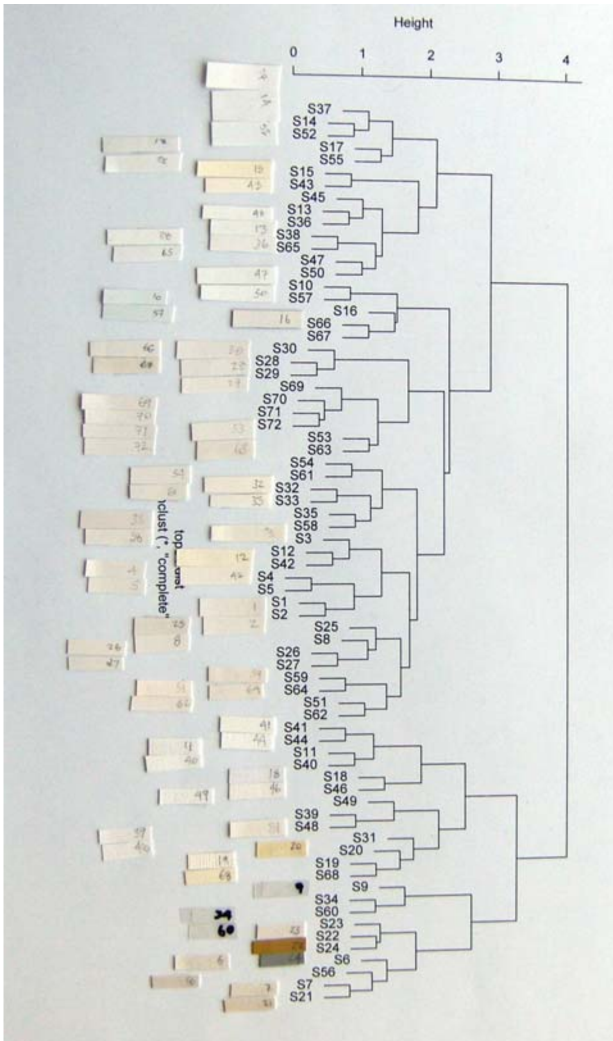


*Fig. 5. Photograph of agglomerative clustering of the 72 print substrates based on incidence of occurrence across the top 64 keywords using the Kendall computation of correlations and visualized as a dendrogram with physical samples glued in place.*

This dendrogram can be interpreted by looking at where the branching occurs — the further to the right, the less similar the terms below are. This clustering shows overall and yet specific trends in natural language usage in the linguistic descriptions of print substrates. There are many specific observations. Over the thickness range considered, subjects categorized substrates into three thicknesses: thin/lightweight, medium and heavy/card/thick/stiff. Likewise, there are separate branches for transparency/film, ivory/yellow/cream/beige and blue/green. There are also branches for gloss terms, texture and metallic/silver/pearl. There are numerous refinements and specific questions which arise from this clustering, but as an initial result for 21 subjects providing anchored unconstrained descriptions of samples, it is also an encouraging result.

Given this matrix of keywords and samples it is also possible to cluster the samples by the patterns of keyword usage. That is, which samples share the greatest number of keywords as used by the subjects. This is simply a rotation of the matrix used to derive Fig. 4. The frequency of usage was again converted to correlations using Kendall's $\tau$ and a distance matrix was computed, which is used with an agglomerative clustering algorithm. The resulting dendrogram is shown in Fig. 5. This figure is actually a photograph of the dendrogram with the corresponding samples glued into the corresponding positions. The results are quite encouraging. The top of the dendrogram has thinner plain papers, including two types of synthetic paper from two different vendors side-by-side. The middle of the dendrogram has thicker semi-gloss substrates followed by glossier substrates. Lower down the dendrogram are the textured media, such as canvas and parchment. Finally appear the transparent media and the substrates described as metallic or with the stem pearl. The clustering of the samples by keyword shown in Fig. 5 provides a useful alternative to the current default alphabetization by vendor. Furthermore this clustering in combination with the keyword clustering provides a natural language, vendor-neutral, more intuitive means to describe these clusters.

## Rehabilitation of Nominal Scales

The use of anchored elicitation of unconstrained text descriptions of visual and tactile properties of print substrates is perhaps a step toward the rehabilitation of nominal scales for investigating perceptual categorization. In 1875 in Psychometric Experiments, Sir Francis Galton, father of the standard deviation and psychometrics, stated: "…until the phenomena of any branch of knowledge have been subjected to measurement and numbers, it cannot assume the status of dignity of a science". Over 100 years later in 1984, Bartleson and Grum[20] noted that nominal scales "merely uses numbers instead of names(sic) to distinguish among members of a group." Is it really the case that nominal scaling is merely[*] about using words and not worthy of the dignity of a science? Furthermore, empirical linguistic techniques, such as corpus linguistics has also faced criticism as being no more than mere[†] lists. We propose that the use of physical anchors provides a specific context for the corpora of categories, that the lack of constraints yields lexical richness and that the use of larger numbers of samples and observers yields a useful collective hierarchy of categories and samples.

We have thus made use of the quantitative techniques of corpus linguistics in combination with the rigor of laboratory perceptual experimental techniques to create a nominal scaling technique. This approach can be contrasted with conventional

---

[*] Figure 2 on page 341 of this reference also inexplicably shows the example ratio scale as being the longest scale. The ratio scale is almost twice as long as the ordinal scale and over 20% longer than the interval scale.

[†] Specifically, Chomsky has objected that "the corpus, if natural, will be so wildly skewed that the description would be no more than a mere list." Chomsky, N., *Aspects of the Theory of Syntax*, Cambridge, MA: MIT Press, p. 159 (1965).

psychometric techniques for the derivation of ratio scales. For example use of a paired-comparison task for 72 samples would require 3,000 judgments per observer per "-ness".[22] Alternatively, use of triadic judgments for 72 samples would require 59,640 triads. There are ways to reduce the sample size in both cases, but it is not clear if the sample size could be reduced sufficiently to fit into a tractable amount of time for volunteer participants. The paired-comparison task also has the limitation that it requires expert definition of the "-ness" such as "glossiness" but does not really address if or how "glossiness" relates to "semi-glossiness" and "matte-ness" other than to perhaps create a scaling task for every term. In contrast, the triadic judgments can be used to avoid this problem but the results require expert interpretation of the resulting scales and there is no corresponding natural language interpretation. In comparison, the results shown here are based on a simple, direct task and result in direct quantitative measures, such as distance metrics, which can be used to cluster keywords by sample, to cluster samples by keyword and in conjunction with the physical measurement data from the introduction can be used for additional applications, such as other multivariate analysis techniques and classification.

The results shown in Fig. 5 provide a global and yet specific view of the visual and tactile properties of commercial print substrates. It has allowed us to efficiently and intuitively navigate print substrate selection in a way that an alphabetical or vendor by vendor listing by application does not. Likewise the results shown in Fig. 6 have allowed us to select terms and communicate properties of print substrates in a way that is vendor neutral and systematic. These terms can also be correlated with the measurement data shown in Figs. 1 and 2 to yield estimates of category measurements based on measurements.

## References


1. N. Ohta and M. Rosen, Color: Desktop Printer Technology, CRC Press, Boca Raton, pp. 237-268 (2006).
2. J. Ferwerda, F. Pellacini and D. P. Greenberg, "A Psychophysically-Based Model of Surface Gloss Perception", Proceedings of SPIE, Human Vision and Electronic Imaging XIII, v. 6806, pp. 291-301 (2001).
3. G. Obein, K. Knoblauch and F. Vienot, "Difference Scaling of gloss: Non-linearity, binocularity, and constancy", Journal of Vision, 4, 711-720 (2004).
4. I. Motoyoshi, S. Nishida, L. Sharan and E.H. Adelson, "Image statistics and the perception of surface qualities", Nature, 447, pp. 206-209 (2007).
5. ASTM D 523-89, Standard Test Method for Specular Gloss, (1989).
6. TAPPI T 480 om-92, Specular Gloss of Paper and Paperboard at 75 Degrees, TAPPI Press, Atlanta (1998).
7. TAPPI T 653 om-98, Specular Gloss of Paper and Paperboard at 20 Degrees, TAPPI Press, Atlanta (1998).
8. R. Bala and R. Eschbach, "Substrate Fluoresence: Bane or Boon?", Proc. 15th IS&T/SID Color Imaging Conference, pp. 12-17 (2007).
9. B. Gamm, "The Characteristics of Optical Brightening Agent Fluorescence Emission and How they Relate to Methods for UV-cut measurement", Proc. ISCC Special Topics Meeting: Black and White, pp. 26-28 (2008).
10. ASTM D 589-97, Standard Test Method for Opacity of Paper (15˚/Diffuse Illuminant A, 89% Reflectance Backing and Paper Backing), (1997).
11. TAPPI T 425 om-96, Opacity of Paper (15˚/Diffuse Illuminant A, 89% Reflectance Backing and Paper Backing), (1998).
12. T.E. Conners and S. Banerjee eds., Surface Analysis of Paper, CRC Press, Boca Raton, p. 80, (1995).
13. Y.X. Ho, M. S. Landy and L. T. Maloney, "How Direction of Illumination Affects Visually Perceived Surface Roughness", Journal of Vision, v. 6, pp. 634-648 (2006).
14. E. Lehtinen, Pigment Coating and Surface Sizing of Paper, Book 11, Gummerus Printing, Jyväskylä, Finland, p. 728 (2000).
15. J. H. Maindonald, Using R for Data Analysis and Graphics: Introduction, Code and Commentary, (2008).
16. ASTM D 985-97, Standard Test Method for Brightness of Pulp, Paper and Paperboard (1997).
17. R. H. Baayen, Analyzing Linguistic Data: A Practical Introduction to Statistics Using R, Cambridge University Press, (2008).
18. T. McEnery and A. Wilson, Corpus Linguistics, Edinburgh University Press (1996).
19. A.M. McEnery, R.Z. Xiao, and Y. Tono, Corpus-based Language Studies: An advanced resource book, Routledge Applied Linguistics Series . London, (2005).
20. C.J. Bartleson and F. Grum eds., Optical Radiation Measurements, Volume 5 Visual Measurements, Academic Press, p. 341 (1984).
21. P. G. Engeldrum, Psychometric Scaling: A Toolkit for Imaging Systems Development, IMCOTEK Press, Winchester, MA (2000).


## Author Biographies

*Nathan Moroney is a principal scientist at Hewlett-Packard Laboratories in Palo Alto, California. Previously, he worked for the Barcelona division of Hewlett-Packard and at the RIT Research Corporation. He has a Masters Degree in Color Science from the Munsell Color Science Laboratory of RIT and a Bachelors degree in color science from the Philadelphia University. His research interests span various color imaging technologies and machine learning techniques. He is a member of the IS&T and the ISCC. Nathan was the technical chair for CIE Technical Committee 8-01, which developed the CIECAM02 color appearance model and created the online color thesaurus.*

*Giordano Bruno Beretta is a technical worker in the Print Production Automation Lab at Hewlett-Packard. He did his graduate work in computational geometry at ETH, before joining Xerox PARC in 1984. At Xerox he has worked on color reproduction and printing, implemented color and imaging standards, and invented color design tools. After working in strategic planning and intellectual property management, and becoming the Technical Advisor for Color at Canon, he joined HP, where he was one of the implementors of a color facsimile machine and the digital sender; he is currently working on printer drivers. He was co-chair of the Electronic Imaging 2000, 2004 Symposia, 1997-1999 EI Color Imaging Conferences, and 2000-2002 EI Internet Imaging Conferences.*