An Analysis of the Factors Influencing Paper Selection for Books of Reproduced Fine Art Printed on Digital Presses

Brian M. Gamm, Franziska Frey and Susan Farnand; Rochester Institute of Technology; Rochester, NY

Abstract

Toner-based digital presses are now capable of matching offset lithographic presses in image and print quality. Current trends show increased interest in printing fine art books on digital presses. It is necessary to understand the extent to which digital printing systems are capable of accurately rendering fine-art reproductions. This research analyzed paper properties that maximize image quality and preference for digitally printed fine art reproductions. Four images, representing four art media, were printed on twelve papers using two digital presses. The twelve papers represented different combinations of color, print-show-through, roughness and gloss. A psychophysical experiment was conducted in which observers ranked the twelve papers for each image on the basis of image quality, color rendering quality, and surface appearance quality. The results were analyzed and a model was developed to predict the probability that a paper was ranked in the top three. Paper color (coolness), basis weight, roughness, and gloss were model parameters. Unlike gloss, roughness, and print-show-through, there was no previous metric for quantifying coolness. Therefore, an additional experiment was conducted to develop a model to predict the perception of coolness using colorimetry. An alternative model was also developed that included parameters such as caliper, print gloss, line raggedness, and dot circularity. The resulting models allowed for the optimization of paper parameters that maximize the probability a paper will produce preferred and high quality images.

Introduction

Books of reproduced fine art are typically printed using offset lithography. Digital presses have only recently become able to produce images of equal quality to offset lithography. The advent of variable data printing and the variety of paper grades available makes digital printing increasingly valuable for short-run and print-on-demand workflows [1]. The predominant issue that arises when comparing offset lithographic and digital presses is that of image and print quality [9, 10]. The earliest digital presses, primarily used for business graphics, lacked the image quality to compete with offset lithography in the reproduction of images. Digital press manufacturers continuously improved their devices and are now able to contend with the print and image quality offered by offset lithography.

Digital presses have been embraced by businesses for the production of marketing and promotional materials, direct mail, transactional and business communications, and on-demand color books [2]. The print-on-demand (POD) capabilities of digital presses have opened the doors for companies, such as lulu.com, to provide print-on-demand services to consumers interested in low-cost self-publishing. However, books of fine art require a higher level of care in printing and collaboration between the artist, publisher and printer, than is offered by POD companies.

Little work has been published, to this authors knowledge, on the use of digital presses for the reproduction of fine art. This research explored a facet of fine art reproduction relating to the selection of substrates and image quality properties of substrates used in digitally printed fine art books. A psychophysical experiment was designed to determine which paper properties maximize image quality for fine art reproduction, and, in the process, demonstrate how statistical design could be used with Engeldrum's Image Quality Circle to create an effective experiment [3, 4, 5, 6].

Methodology

This study was designed using the principles of Engeldrum's Image Quality circle which provides a method for modeling human response based upon physical measurement [5]. Observers in this experiment first ranked the images on the basis of image quality then ranked them on the basis of color rendering quality and surface appearance quality. The latter two rankings were akin to Engeldrum's Customer Perceptions, or 'nesses,' while image quality was akin to Customer Image Quality Rating. The sections below describe the processes by which samples were selected, prepared, and printed, followed by the psychophysical design and practical methods for running the experiment.

Sample Selection

Four factors were identified by which paper could be selected using visual and tactual methods: roughness, gloss, print-show-through (PST), and color. All were easily distinguishable using touch or sight. Each factor was categorized into two levels to satisfy a 2^k factorial sampling structure: rough and smooth for roughness, high and low for gloss, high and low for PST, and warm and cool for color.

Sixteen papers where needed to fulfill a full factorial sampling structure. However, it was determined, based upon conversations with representatives of paper manufacturers and distributors, that papers containing high roughness and high gloss were not manufactured. Thus, the four elements of the full factorial design containing rough and glossy papers were removed. Table 1 shows the 12 combinations of color, PST, roughness, and gloss that where included in the study.

The sample selection as originally designed to keep basis weight constant. However, there was not enough noticeable variation in PST among papers of the same basis weight. Therefore, basis weight was included as a confounding factor with PST. All papers with high PST where 80lb text weight and all low PST papers were 100lb text weight. One paper was selected to fulfill each of the twelve criteria listed in Table 1. All but one paper was

Table 1. Sample selection design for the 12 paper samples included in this study.

Paper Code	Color	PST	Roughness	Gloss
A	cool	low	smooth	low
B	cool	high	smooth	low
C	cool	low	rough	low
D	cool	high	rough	low
E	cool	low	smooth	high
F	cool	high	smooth	high
G	warm	low	smooth	low
H	warm	high	smooth	low
I	warm	low	rough	low
J	warm	high	rough	low
K	warm	low	smooth	high
L	warm	high	smooth	high

known to have been used in fine art books or was of interest to the fine art reproduction community.

Test Targets and Sample Design

This experiment required two press runs to produce test targets and experiment samples.The first press run produced IT8.7/4 profiling targets and targets from which print quality attributes were measured. Two test targets were printed during the first press run to provide a variety of different measurements. Several predictors were chosen that the experimenters felt would illuminate important aspects of print quality: 0.1mm dot circularity, line raggedness, 40% print mottle, print gloss (100% CMY), gamut volume, and solid ink density. Two targets were printed from which these measurements were made: a target from Quality Engineering and Associates (QEA) designed for use with QEA's Image Analysis System Lab (IASLab^{TM}) software [7], and a custom target containing 100% and 40% CMY solid area patches.

The experimental samples were printed during the second press run. Figure 1 shows the four images used in the experiment. Each image represented a different art medium: an aquatint print, an oil painting, a watercolor painting, and a sepia platinotype photograph from the archives of the RIT Image Permanance Institute, corresponding to Figures 1a-d, respectively.

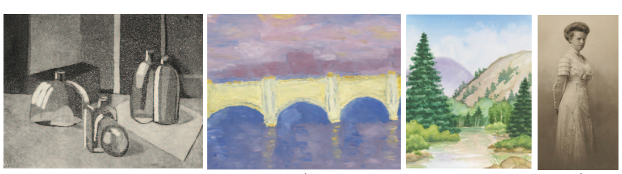

Figure 1. Sample Images

A sample book was created (Figure 2) using four repeated prints of the images shown in Figure 1. A block of Lorum Ipsum text backed up each image on the reverse side of the sheet. The book's dimensions were 8x10 in. and was organized in four sections, one for each image. Each section contained four replicates of its respective image followed by ten unprinted sheets. The four image replicates were available to replace pages that succumbed to damage. Ten unprinted sheets in each section added bulk to the sample books. The cover was designed as a wrap-around and used a cover-weight paper different from those in the books. The book was bound using a 1/2 inch spiral bind. A label was placed on the cover of each book containing codes signifying press and paper. Figure 2 shows three views of an example book.

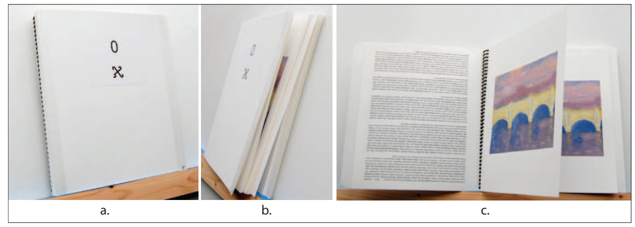

Figure 2. Sample Book Images

One book was printed for each press and paper combination, resulting in 24 books. The books were stored in a file cabinet in the experiment room to avoid excess light exposure when not in use.

Printing Process

The experiment required that both a liquid toner and a dry toner press be included. Each press run included a run on an HP Indigo 7000, a liquid toner press, and a Kodak NexPress S3000, a dry toner press. The presses were calibrated prior to each run. Following calibration, the measurement and profiling test targets were printed on each press. Four IT8.7/4 test targets were printed for each substrate then measured using an X-Rite iSis XL with Measure Tool 5.8.10. ProfileMaker 5.8.10 was used to create the profiles. An early binding workflow was used; however, an adjustment to the workflow was required for the HP Indigo RIP's built-in color management workflow. The ICC profiles embedded within the original PDF files were removed using Enfocus PitStop Pro. The CMYK values were unaffected by this process. The numbers were verified in Acrobat Professional 9 using Enfocus PitStop Pro. The same workflow was used for the Kodak NexPress S3000 to ensure consistency between the two presses.

Psychophysical Experiment

The experiment took place at the Munsell Color Science Laboratory (MCSL) at the Rochester Institute of Technology. Observers were first led into a room with a light booth containing the four original works of art under tungsten illumination. Two additional works were added to the booth to make the experience more akin to a gallery experience. The experimenter then led observer to a D50 viewing room in another area of the lab where the experiment was conducted. The experiment was designed in four sections: Image Quality, Color Rendering Quality, Surface Appearance Quality, and Preference. For the Image Quality Experiment, observers were instructed to judge the images using any criteria. For the Color Rendering Quality Experiment, observers were instructed to consider only factors relating to color, and for the Surface Appearance Quality Experiment, observers were instructed to consider only factors relating to the paper surface and print quality. Observers ranked the books according to the criteria outlined for each section. The books were displayed on two wood boards clamped to a Commando XX tilting table (see Figure 3a-c). All four images were ranked independently in the Image Quality, Color Rendering Quality, and Surface Appearance

©2011 Society for Imaging Science and Technology

Quality sections. Observers were instructed to use the provided space however they felt was most efficient for making their judgments and were allowed to handle the books as they saw fit. Observers handed books to the experimenter in the order of image quality. The experimenter recorded this order. Figure 3b shows an observer conducting the experiment.

Figure 3. *Sample Book Setup*

Experiment Observers

One hundred sixteen observers, mostly RIT students and faculty, participated in this experiment. Sixty-one percent of observers were female and 39% were male. Fifty percent of observers were between 18 and 22 years old. While this age range is representative of the undergraduate student population, not all undergraduate students were between 18 and 22 years old. A wide range of fields was also represented in the observer population. This included undergraduate and graduate fields of study in addition to career fields because such a large percentage of students participated in the study (91% students). Twenty nine percent of observers were 18 or 19 years old. An additional 8% of students were directly affiliated with the RIT Center for Imaging Science.

Physical Measurements

Physical measurements of the papers were made for the four sample selection factors. In addition, several other factors not included in the paper selection process were also measured. Color was quantified using a metric called Coolness, derived from a separate experiment that quantified the perception of paper coolness using colorimetry. This experiment will be described in future work. Roughness, recorded in micrometers of air flow across the paper surface, was measured using the ISO 8791-4:2007, Print Surf method with a Testing Machines Inc. Parker Print Surf device. Paper and print gloss were measured using the 60 degree method, as described by ASTM D523-08 and ISO 2813:1994, using a Color Control Systems ETB-0833 Glossmeter. Print-show-through was a metric developed at International Paper, and is calculated using Eq. 1,

$$Percent\ PST = \frac{\Delta L^*_{backside}}{\Delta L^*_{frontside}} \times 100 \qquad (1)$$

where, $\Delta L^*_{backside}$ is the difference between L^* on the backside of a printed 100% CMY patch and paper white, and $\Delta L^*_{frontside}$ is the difference between L^* of a printed 100% CMY patch and paper white. The percent PST metric is the best available metric relating to the visual experience of perceiving duplex printed samples. The automated measurement cabability of IASLab^{TM}was used to measure 40% mottle, line raggedness, and dot circularity from a single scanned test-target image. Caliper, recorded in micrometers, was measured for each paper stock. Solid ink density was measured using an X-Rite i1 Pro and Gamut Volume was calculated using Chromix ColorThink 3.0 Pro and measurements from the IT8.7/4.

Results and Discussion

The Image Quality Circle upon which the psychophysical models were based is shown in Figure 4. The Technology Variables component is shown for reference, but was not included in this experiment. The Physical Image Parameters include the sample selection parameters: coolness, PST, roughness, and gloss, and the additional print quality parameters 40% print mottle and line raggedness. The Customer Perceptions included both Color Rendering Quality and Surface Appearance Quality.

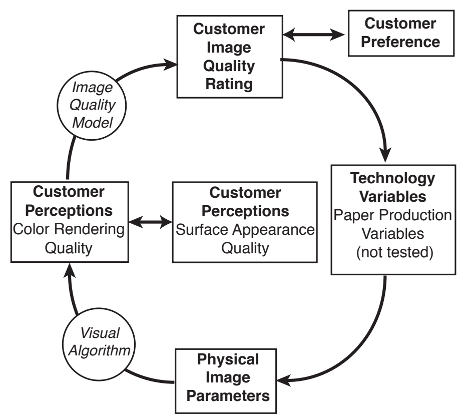

Figure 4. *The Image Quality Circle on which the psychophysical models were based.*

Predictor Selection and Response Definition

The complete set of Physical Image Parameter predictors included: coolness, PST, roughness, gloss, caliper, mottle, dot circularity, line raggedness, print gloss, print contrast, and solid ink density. The predictors were analyzed for multicollinearity and it was found that several predictors were linearly dependent. Multicollinearity can result in larger than expected variances and least square estimates [11]. Thus, several predictors were removed from the set. The first four predictors, those used for sample selection, remained, along with 40% print mottle and line raggedness. A scatterplot of the final predictor set is shown in Figure 5.

Customer Perception scales were generated from the ranking data of individual observers. The ranking data from all observers was combined for each image and each press. Several methods are available for transforming rank data to cumulative probabilities and then to standard normal scores using the inverse normal distribution. The Hazen method was developed for use in water quality analysis applications and found to coincide well with parametric methods used in that field, also works well with psychophysical rank data [8]. Rank data from each observer for the Color Rendering Quality and Surface Appearance Quality experiments were transformed to Hazen cumulative probabilities using Eq. 2,

$$F_{hazen}(x_i) = \frac{n + \frac{1}{2} - x_i}{n} \tag{2}$$

where, n is the number of samples, 12 in this case, and x_i is the i^{th} rank. Finally, the Hazen cumulative probabilities were transformed to standard normal scores, z-scores, using Eq. 3.

$$u = \Phi^{-1}(F_{hazen}(x_i)) = \Phi^{-1}\left(\frac{n + \frac{1}{2} - x_i}{n}\right) \tag{3}$$

The standard normal scores were averaged for each book across observers for each image and press combination. Thus, eight sets of mean standard normal scores (four images printed on two presses) were collected for the Color Rendering Quality experiment and for the Surface Appearance Quality experiment.

Unlike Customer Perceptions, Customer Image Quality Rating describes what an observer likes and does not like. Analyzing the first three rank positions provided the most relevant picture of what observers liked. Thus, instead of transforming the rank data to a standard normal distribution, the number of times a particular book was selected in the first, second, or third ranking position was tallied. No dependence on image or press was found.

Under the null hypothesis of random ranking the probability of selecting a book in any one position was the same, and thus equal to 0.083. Binomially distributed data may approximate a normal distribution subject to the constraints, $np \geq 10$ *and* $n(1-p) \geq 10$, where n is the number of observations, in this case 58 observers for each press, and p is the probability a book will be selected either first, second or third. The Image Quality and Preference count data did satisfy the above constraints and was transformed, using Eq. 4 to standard normal scores. Note that the empirical probability of selecting a book either first, second or third, is equal to $3\ x\ p = 0.25$.

$$Count_{stand} = \frac{Count - np}{\sqrt{np(1-p)}} \tag{4}$$

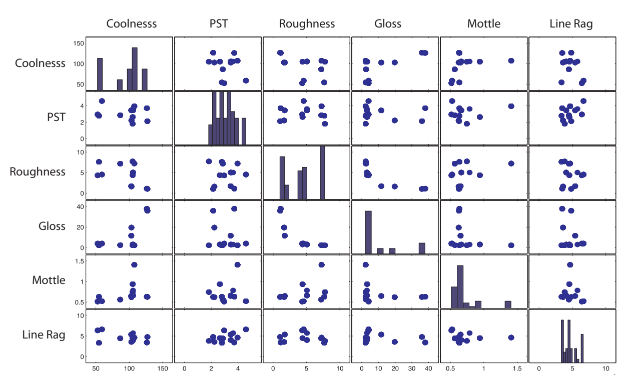

Figure 5. *Scatterplot of the final set of Physical Image Parameters.*

Visual Algorithms

Four different visual algorithms were analyzed: Color Rendering Quality predicted using the four sample selection parameters, Color Rendering Quality predicted using the expanded set of parameters, Surface Appearance Quality predicted using the four sample selection parameters, and Surface Appearance Quality predicted using the expanded set of parameters. Stepwise multiple regression models were computed for each Visual Algorithm using the MATLAB® function *stepwisefit*. The function determined the set of parameters that minimized the RMS error between the predicted and actual Customer Perceptions. The result was that both Visual Algorithms predicting Color Rendering Quality were the same (adj. $R^2 = 0.85$), as were both Visual Algorithms predicting Surface Appearance Quality (adj. $R^2 = 0.30$). In addition, Color Rendering Quality and Surface Appearance Quality were predicted using the same three Physical Image Parameters, coolness, roughness, and gloss, although with different regression coefficients and levels of success (illustrated by the adj. R^2 values). The regression statistics are shown in Table 2.

Table 2. Regression statistics for the Color Rendering Quality and Surface Appearance Quality Visual Algorithms.

Color Rendering Quality				
	Intercept	**Coolness**	**Roughness**	**Gloss**
B	-0.979	0.020	-0.141	-0.022
p-val		< 0.001	< 0.001	< 0.001
Surface Appearance Quality				
	Intercept	**Coolness**	**Roughness**	**Gloss**
B	0.090	0.004	-0.081	-0.009
p-val		0.003	< 0.001	0.027

Plots of the four design predictors and two expanded predictors against the Color Rendering Quality Hazen z-scores are shown in Figure 6. The respective plot for Surface Appearance Quality is only slightly different. Residuals were analyzed for the two Visual Algorithms. The three significant predictors, coolness, roughness, and gloss, are shown in red. Two outliers were previously removed from the data due to excessive abrasion on one of the sample books. Assumptions of residual normality, equal variance and randomness as a function of experiment order were evaluated using standardized residual QQ plots, plots of standardized residuals versus model fits, and plots of standardized residuals against experiment order. The plots suggest that all assumptions were satisfied.

Image Quality Model

According the Image Quality Circle, Customer Image Quality Preference is predicted directly from Customer Perceptions. The model used to make that prediction is called an Image Quality Model. Customer Image Quality Rating was measured by the Image Quality stage of the psychophysical experiment. It may be interpreted as a measure of behavior, describing choices people make based upon likes and dislikes, while Customer Perceptions are measurements of subconscious elements of decision making, forcing observers to focus on elements of decision making of which they would not normally be attentive.

Multicollinearity was analyzed by plotting the fitted values of Color Rendering Quality against the fitted values of Surface

©2011 Society for Imaging Science and Technology

Appearance Quality modeled by the Visual Algorithms, shown in Figure 7a. The plot suggests a strong correlation between Color Rendering Quality and Surface Appearance Quality, which indeed was the case ($r = 0.92$, $p << 0.001$). Therefore, it was necessary to choose either the Color Rendering Quality or Surface Appearance Quality model. Plots of Standardized Image Quality Counts against Color Rendering Quality and Surface Appearance Quality are shown in Figures 7b and 7c, respectively.

Color Rendering Quality was used as the predictor for Image Quality because the Color Rendering Quality was tested before Surface Appearance Quality during the psychophysical experiment. Thus, it is highly likely that observers were influenced by having previously analyzed color quality when judging surface appearance quality.

Standardized residual plots for the linear fit between Color Rendering Quality fits and Standardized Image Quality Counts are shown in Figure 8. The normal probability plot in Figure 8a suggests the residuals are not normally distributed. This may be attributed to the possibility that a linear fit is not ideal or that there are unattributed factors at work. However, further research must be conducted to determine whether another model would be more suitable and whether there is a psychological basis for a nonlinear model. The remaining residual plots, in Figures 8b and 8c, do not suggest a lack of equal variance or present outliers. The cyclical nature of 8c is due to the repetition of rank data across image and press.

Contour plots are shown in Figure 9 illustrating the change in Color Rendering Quality as a function of coolness, roughness, and gloss. The response is shown as a z-score, and changes positively as a function of coolness, negatively as a function of roughness, and negatively as a function of gloss. Thus, a paper expected to have a high Color Rendering Quality will have high coolness, low roughness, and low gloss. By extension, Surface Appearance Quality and Customer Image Quality Rating are also optimized for papers with high coolness, low roughness and low gloss.

The Image Quality Models were designed to model only specific judgments of quality, and not the preference of observers. It is often assumed that images rated as having the highest quality will be most preferred by observers. Discussions with observers following each of the first three experiments revealed that, for the most part, observers objectively judged quality aspects of the images, rather than incorporating their personal preferences. In the fourth experiment, observers were asked to select their three most favorite books. While the Image Quality Circle did not explicitly model preference, the relationship between Customer Image Quality Rating and Customer Preference was analyzed to determine if Customer Preference could be predicted by Customer Image Quality Rating. A scatterplot is shown in Figure 10 depicting this relationship. There is a significant correlation between Customer Image Quality Rating and Customer Preference ($r = 0.73$, $p << 0.001$), suggesting a strong linear relationship.

Table 3. The two final models for the Image Quality Circle

Model Component	Model
Visual Algorithm	$CQ = -0.9791 + 0.0196 * C - 0.1411 * R - 0.0219 * G$
Image Quality Model	$CIQR = -0.0611 + 3.6880 * CQ$
CQ = Color Rendering Quality, R = Roughness G = Gloss, CIQR = Customer Image Quality Rating	

The psychophysical experiment modeled observers' perceptions of image quality using common print and paper quality metrics and provided both specific results and general commentary on

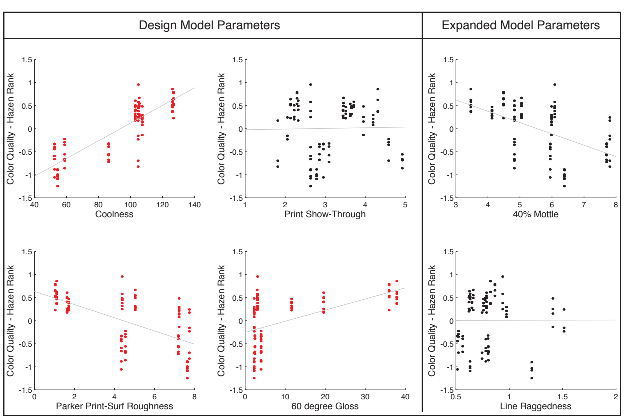

Figure 6. Plots of the six Physical Image Parameters against Color Rendering Quality z-scores.

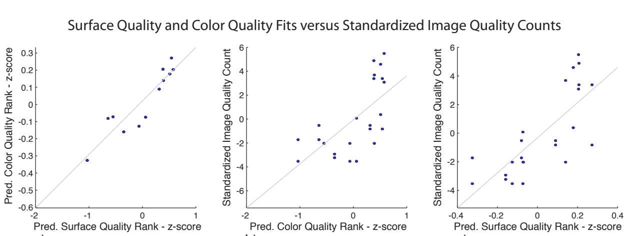

Figure 7. Plots of Color Rendering Quality versus Surface Appearance Quality, and Customer Perception fits versus Customer Perceptions.

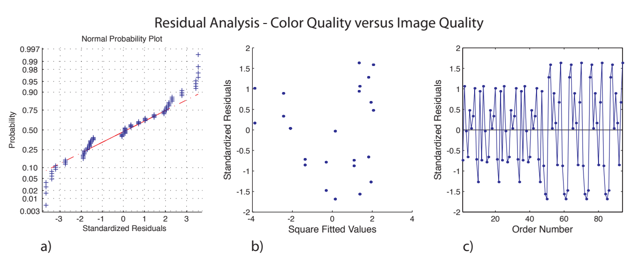

Figure 8. Standardized residual plots for the linear fit between Color Rendering Quality fits and Standardized Image Quality Counts

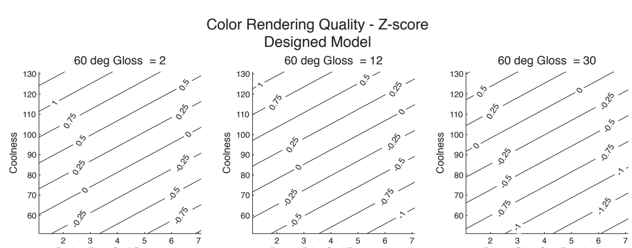

Figure 9. Contour plots illustrating the change in Color Rendering Quality as a function of coolness, roughness, and gloss

the practice of conducting psychophysical experiments. Specifically, a model was developed, in accordance with the Image Quality Circle, to predict the paper properties maximizing the potential a paper is selected for a book. This model is shown in Table 3. Coolness, roughness, and gloss were the significant Physical Image Parameters for both the designed and expanded experiment models. Those predictors, weighted differently, were used in both the Color Rendering Quality and Surface Appearance Quality Visual Algorithms. Color Rendering Quality was chosen as the single customer perception in the Image Quality Model. One can say that choosing a paper with high coolness, low roughness, and low gloss maximizes the chance of producing a high quality image. These data are based upon a small number of papers, though representative of many, printed on two digital presses.

This research explored the use of statistical design in conjunction with the Image Quality Circle to conduct a psychophysical image quality experiment. Most documentation does not cite an exploration of the Image Quality Circle to the extent described in this report. The experiment limited predictors to those that are commonly used for print and paper quality. The significant predictors for the Customer Perception experiments were the same, suggesting that observers were influenced by the same criteria when ranking the books in both experiments. A likely explanation for this disconnect was bias resulting from experiment order. The Image Quality experiment was chosen as the first experiment. However, while judging image quality, observers may have formed biases to the samples which remained throughout the experiment.

The frequency of words and concepts mentioned by observers during the post-experiment interviews was also analyzed. The top-ten most frequently used words by observers following the Image Quality experiment was equally divided between words related to Color Rendering Quality and Surface Appearance Quality. As expected, nine of the top-ten most frequently used words following the Color Rendering Quality and Surface Appearance Quality experiments were related to color and paper surface, respectively. Yet, the linear models predicting both Customer Perception responses were similar. Either the Physical Image Parameters may not have been suitable predictors for this analysis or observers did not rank the images as they indicated in their post-experiment interviews, supporting the claim of observer bias. The solution, in future experiments, would be to randomize the order of the experiments or limit observer participation to a single experiment, ensuring that effects of bias would be minimized.

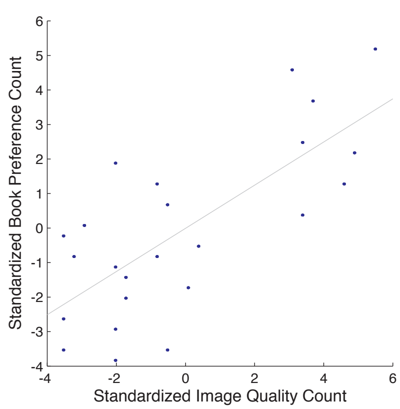

Figure 10. *A scatterplot showing the relationship between Customer Image Quality Rating and Customer Preference*

Conclusion

A psychophysical experiment analyzing the relationship between paper quality metrics, Customer Perceptions, and Customer Image Quality Ratings revealed that Image Quality was optimized using papers with high coolness, low roughness, and low gloss. This study demonstrated the use of statistical design and the Image Quality Circle for the analysis of image quality of printed images. However, the design did not effectively account for observer bias as a result of experimental order. Further experiments will attempt to better control observer learning and bias through randomization of experimental order.

Acknowledgments

The authors would like to thank the Printing Industry Center and the McGhee Chair for their contributions to this research, to Mohawk Paper for their generous paper donation, and to the many observers without whom this research would not have been possible.

References

[1] Howard Vogl, A Survey of Digital Press Manufacturers: Critical Paper Requirements, Printing Industry Center, Rochester, NY, 2008.
[2] Franziska Frey, Henrik Christensen and Nicholas DiSantis, Permanence of Toner on Paper Based on the Life Cycle of Documents, Printing Industry Center, Rochester, NY, 2008.
[3] Peter Engeldrum, Psychometric Scaling: A Toolkit for Imaging Systems Development, Imcotek Press, Winchester, MA, 2001.
[4] Peter Engeldrum, Extending Image Quality Models, Proc. PICS, pg. 65. (2002).
[5] Peter Engeldrum, A Theory of Image Quality: The Image Quality Circle, J. Imaging. Sci. and Technol., 48, 5, pg. 447. (2004).
[6] Peter Engeldrum, A Short Image Quality Taxonomy, J. Imaging. Sci. and Technol., 48, 2, pg. 160. (2004).
[7] Quality Engineering Associates, Inc, IASLab [computer software], Burlington, MA, 2009.
[8] P.R. Hunter, Does calculation of the 95th Percentile of Microbiological Results Offer Any Advantage Over Percentage Exceedence In Determining Compliance with Bathing Water Quality Standards?, Letters in Applied Microbiology, 34, pg. 283. (2002).
[9] Susan Farnand, Minding the Gap: Evaluating the Image Quality of Digital Print Technologies Relative to Traditional Offset Lithography, Printing Industry Center, Rochester, NY, 2008.
[10] Susan Farnand, Further Investigation Into the Image Quality Differences Between Digital Print Technologies and Traditional Offset Lithography, Printing Industry Center, Rochester, NY, 2009.
[11] D.C. Montgomery, E.A. Peck and G.G. Vining, Introduction to Linear Regression Analysis, Ed. 4, John Wiley & Sons, Inc., New York, 2006.

Author Biography

Brian Gamm is currently studying for his Master's degree at the Munsell Color Science Laboratory, at RIT in Rochester, NY. He graduated from RIT in 2009 with a BS in Imaging and Photographic Technology and will begin study for his PhD in Color Science in Fall 2011. He is currently a member of IS&T.

©2011 Society for Imaging Science and Technology