Comparing image quality of print-on-demand books and photobooks from web-based vendors

Jonathan Phillips
Rochester Institute of Technology
Munsell Color Science Laboratory
Chester F. Carlson Center for Imaging Science
Rochester, New York 14623
E-mail: jbp3319@cis.rit.edu

Peter Bajorski
Rochester Institute of Technology
Graduate Statistics Department
Rochester, New York 14623

Peter Burns
Carestream Health Inc.
Rochester, New York 14615

Erin Fredericks
Rochester Institute of Technology
Munsell Color Science Laboratory
Chester F. Carlson Center for Imaging Science
Rochester, New York 14623

Mitchell Rosen
Rochester Institute of Technology
Munsell Color Science Laboratory
Chester F. Carlson Center for Imaging Science
Rochester, New York 14623

Abstract. *Because of the emergence of e-commerce and developments in print engines designed for economical output of very short runs, there are increased business opportunities and consumer options for print-on-demand books and photobooks. The current state of these printing modes allows for direct uploading of book files via the web, printing on nonoffset printers, and distributing by standard parcel or mail delivery services. The goal of this research is to assess the image quality of print-on-demand books and photobooks produced by various Web-based vendors and to identify correlations between psychophysical results and objective metrics. Six vendors were identified for one-off (single-copy) print-on-demand books, and seven vendors were identified for photobooks. Participants rank ordered overall quality of a subset of individual pages from each book, where the pages included text, photographs, or a combination of the two. Observers also reported overall quality ratings and price estimates for the bound books. Objective metrics of color gamut, color accuracy, accuracy of International Color Consortium profile usage, eye-weighted root mean square* L^**, and cascaded modulation transfer acutance were obtained and compared to the observer responses. We introduce some new methods for normalizing data as well as for strengthening the statistical significance of the results. Our approach includes the use of latent mixed-effect models. We found statistically significant correlation with overall image quality and some of the spatial metrics, but correlations between psychophysical results and other objective metrics were weak or nonexistent. Strong correlation was found between psychophysical results of overall quality assessment and estimated price associated with quality. The photobook set of vendors reached higher image-quality ratings than the set of print-on-demand vendors. However, the photobook set had higher image-quality variability. © 2010 SPIE and IS&T.* [DOI: 10.1117/1.3271167]

1 Introduction

Color printing is seeing a tremendous growth rate in the nonoffset digital printing category.[1] Print engines have begun to enable inexpensive output of very short-run documents right down to one-off (single-copy) books that can sell for as little as $10 a unit. The World Wide Web has spawned new business models in which individuals can self-publish print-on-demand art books or books for small or no fee by uploading their content online.[2–4] Web-based publishers have begun to capitalize on this new approach, creating menus and information guides for individuals wishing to print and publish books.[5–10] Some websites even

Paper 09081SSRR received May 18, 2009; revised manuscript received Sep. 23, 2009; accepted for publication Oct. 5, 2009; published online Jan. 8, 2010.

1017-9909/2010/19(1)/011013/17/$25.00 © 2010 SPIE and IS&T.

offer assistance in obtaining ISBN numbers.[11] In addition, Web sites are creating new international social networking environments for book publishers.[5]

There are several categories of book types generally supported by online publishers: typical books, pamphlets, and photobooks. Typical books may come in hard- or softcover, in a selection of sizes with one of several options of binding style, and with a choice of media quality and weight. Pricing is typically quoted per order and normally dependent on the number of pages and copies.

In the category of photobooks, each vendor typically supplies software to use for assembling each page. Many vendor options exist in the marketplace.[12–18] For a given page, photos as well as text can be placed in preassigned positions from a selection of preassigned templates. Because the pages are created one at a time, setting up a photobook can be very time consuming. In response, current research is studying ways of simplifying and automating the process of selecting and arranging photos for the consumer.[19–22]

The goal of the study described here was to provide baseline data of the current state of one-off typical books and photobooks in terms of delivered image quality and price expectations. The first experiment was concerned with typical print-on-demand books. An approximately 90-page book (with most pages one sided) was designed and, through Web interfaces, was fed to each of six on-line vendors for softcover publication. The one-off printed and perfect bound books were delivered through typical services, such as UPS, Federal Express, and U.S. Mail. Psychophysical studies were run to gain understanding of how perceived image quality influenced price estimates and corresponded to physical measurements. Observers evaluated individual pages that had been removed from the books and also looked at the books in their intended format as bound books. For the second experiment, photobooks were investigated. Seven vendors were selected, and a hardcopy photobook was made with the software from each vendor. Similar to the first experiment, psychophysical studies were run in order to compare physical measurements to the overall image-quality assessment and to determine the influence of image quality on price estimation.

In Section 4, we provide results and details of our analyses. We use some innovative ways to analyze quality-assessment data, including some new methods for strengthening the statistical significance of the results. Our approach includes the use of latent mixed-effect models.

2 Book Preparation

2.1 *Experiment 1: Print-on-Demand Books Preparation*

2.1.1 *Print-on-demand vendors*

Table 1 contains the printer information for the six vendors selected for the first experiment. At the time of selection during the summer of 2007, these vendors were among the few identified by research as online options in the United States that offered one-off printing for $<\$150$ per book. One exception was vendor A that required a minimum purchase of two books.

Table 1 Printer information for the six selected vendors offering one-off print-on-demand books. Information was obtained from the vendor website or direct communication with the vendor.

Vendor	Printer
A	Xerox DocuColor DC250
B	Xerox iGen
C	Xerox iGen
D	IBM InfoColor 70
E	HP Indigo 1000
F	HP Indigo 5000/5500

2.1.2 *Print-on-demand content*

Book content was collected for these experiments with the goal of assessing pictorial and text image quality as well as to determine the influence each has in determining overall image quality for print-on-demand applications. In addition, the pages were designed to allow analysis of accuracy of International Color Consortium (ICC) profile usage, color gamut volume, line quality, overall contrast, and print uniformity. This was accomplished by creating pages with text, photos, and combinations of these. There were also pages with various types of evaluation targets that enabled the color and spatial analysis. Pages included the IT8.7/3 Extended Ink Value Data Set,[23] a simulated Macbeth ColorChecker created from BabelColor sRGB code values,[24] ISO/CD 12640-3 CIELAB SCID images,[25] as well as pages of photos, text, and targets made specifically for this set of experiments. Photos included various skin tones and other key memory colors, such as sky, foliage, and fruit. Replicates of pages were included in the book in order to allow for removal of single pages to be used in the psychophysical testing. The final book content was 91 page sides that, for most pages, only had one side with printed content.

2.1.3 *Print-on-demand color encoding*

To assess various processing paths, color encodings in the book included RGB, CMYK, K only, and 16-bit CIELab. Some components were tagged with ICC profiles while others were purposely untagged. Representative pages are found in Fig. 1.

2.1.4 *Print-on-demand profiles*

ICC profile interpretation ranging from correct to incorrect to complete disregard can cause significant differences in color printing output. In order to assess the profile interpretation accuracy, the following embedded ICC profiles were attached to certain objects within the book: sRGB, Adobe RGB, sYCC, e-sRGB, and GBR, a pathological profile.[26] Figure 2 illustrates a target comprised of four subimages designed for quick visual assessment of how well a system handles various types of version 2 and version 4 ICC profiles.[27] Adobe RGB and GBR are version 2 while e-sRGB and sYCC profiles are version 4. Each quadrant

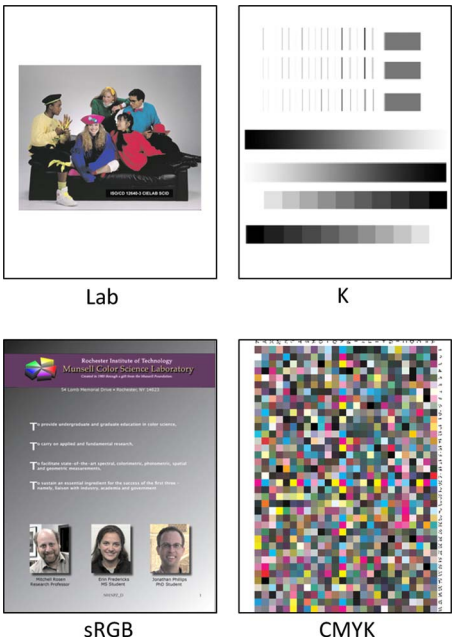


Fig. 1 Representative book pages with indicated color encoding.

contains an embedded profile and associated code values such that the target looks like a single image when printed correctly. When profile misinterpretation or disregard takes place, the quadrants should appear distinct. See Fig. 3 for an example where the profiles have been ignored. Details on how the various vendors handled profiles in the print-on-demand books can be found in Section 4.1.4.

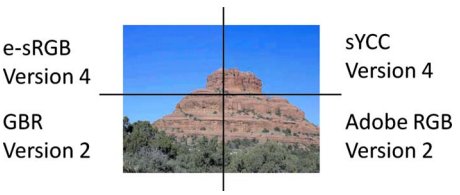


Fig. 2 ICC profiling target with four quadrants, each with identified embedded profile and profile type. Image shown with correct ICC profile interpretation. (Used with permission.)

Fig. 3 Example print of ICC profiling target from Fig. 2 when interpreted incorrectly as sRGB for each quadrant. (Used with permission.)

In addition to images with profiles, the book also contained objects without embedded profiles. This allowed us to see how untagged CMYK, K, and RGB elements were handled.

2.1.5 *Print-on-demand preparation*

The files for the print-on-demand books were constructed using Adobe InDesign CS2/CS3, then exported as PDF files (PDF 1.7). For color and gray-scale images, the export settings turned off downsampling and JPEG compression was set at the maximum image-quality setting. Confirmation of correct profiling was performed with a trial version of an Adobe Acrobat plug-in [Quite Revealing 1.8a (EN), Quite Software]. Because of the complex nature of embedding profiles and exporting to PDF with InDesign, two sections of the book were separately exported—the main section and a three-page subsection. The main section was exported using settings that preserved embedded ICC profiles and untagged K and CMYK sections. The subsection was exported using a process that did not interfere with untagged RGB images. Pages from both sections were combined in Adobe Acrobat 8.1.1 Professional prior to uploading to each vendor's Web site. See Fig. 4 for a flowchart of the process described above. The raster content was set at either 300 or 600 dpi, depending of the spatial needs of the page content. Vector graphics were maintained.

2.1.6 *Print-on-demand printing*

Each Web site supplied instructions for uploading the book PDF files. The particular book files in this experiment were ~90 MB. Books were printed in late 2007.

2.2 *Experiment 2: Photobook Preparation*

2.2.1 *Photobook vendors*

Seven key photobook vendors in the United States were selected for experiment 2, designated, respectively, vendor G, H, I, J, K, L, and M in this paper. Each vendor supplied downloadable software to create the pages of the photobooks. Printer details were not successfully obtained.

2.2.2 *Photobook content*

The content of the photobooks was as consistent as possible with the print-on-demand content. However, page content had to be assembled individually rather than uploaded as a prepared PDF document as in experiment 1. Exceptions are described below.

2.2.3 *Photobook color encoding*

Photobook offerings are aimed at the consumer photographic market. Therefore, the default assumption of the software was that the images were RGB in format. To match this assumption, several image types from the print-on-demand book design needed to be converted to RGB values so that the same images could be used in all photobooks. Specifically, the CMYK and 16-bit Lab files were converted to RGB via Adobe Photoshop (CS3 Extended, Version 10.0.1) using the North America General Purpose 2 color settings, which assume sRGB and version 2 specifications web offset publications. The conversion utilized the Adobe engine and relative colorimetric intent with black point compensation and dither. In addition, the gray-scale

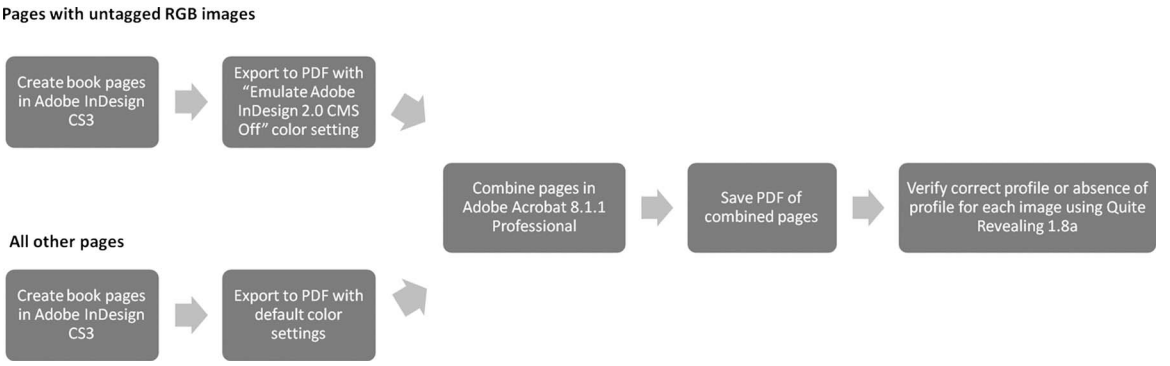


Fig. 4 Flowchart of process to create tagged (profiled) and untagged (unprofiled) pages of book.

images were converted to RGB mode, also in Photoshop, using dot gain 20% for vendors I and M because the software would not accept K.

2.2.4 *Photobook profiles*

Embedded profiles were maintained in files where present. For vendor I, the images with e-sRGB and sYCC were eliminated from the book because its software would not allow images with these embedded profiles to be uploaded. Because the offending profiles were both ICC version 4 (see Fig. 2), it is clear that as of the time of accessing vendor I, the only accepted tagged images were those with version 2 profiles. Further discussion of how the various vendors treated profiles in the photobooks can be found in Section 4.2.4.

2.2.5 *Photobook preparation*

The differences in photobook software led to several book-to-book differences, such as in the positions and sizes of photos on compilation pages. In addition, some books used the "fill" option, which cropped images into a page template to completely cover the predefined region with image content while other books used the "fit" option, which ensured that all image content appeared in the predefined region (leaving white space if the aspect ratio of the image and template differed). Vendor K's photobook was limited to 80 pages total; thus, blank pages were eliminated to meet this requirement. Vendor L would not accept the 600-dpi images due to file-size constraints. The images were converted to 300 dpi via Photoshop for this vendor.

2.2.6 *Photobook printing*

Photobooks were created with each vendor's software and then uploaded for printing in early 2008.

3 Psychophysical Testing

The psychophysical testing was divided into two parts: evaluation of individual pages and evaluation of bound books. These individual pages as well as target pages were removed from the book prior to the bound-book evaluation. Most pages were one sided. Thus, the final number of page sides evaluated in each intact print-on-demand book was reduced to 40 pages plus two full-color covers. The final number of pages in the photobooks was reduced to 42 pages.

3.1 *Experiment 1: Print-on-Demand Book Psychophysical Testing*

3.1.1 *Print-on-demand individual pages*

Ten individual pages were selected for quality evaluation. Some had photos only ($n=6$), some had text only ($n=1$), and some had a combination of text and photos or charts ($n=3$). For a set of a given page from each vendor, participants were asked to order the pages for overall image quality. The vendors were then assigned a quality score (QS) from 1 to 6 (1=worst, 6=best) for that page. This was repeated for all pages.

3.1.2 *Print-on-demand perfect-bound books*

Observers rated overall image quality of the print-on-demand books on a scale from 1 to 5 (1=very low satisfaction, 5=very high satisfaction) (referred to as "ratings" here). The observers were also asked how much they would be willing to pay for this quality of book as a memento of the participant's vacation (called "vacation price" here) and as lecture notes to be given out during a presentation (called "notes price"). Book evaluations were performed in a single-stimulus fashion.

3.1.3 *Experimental details*

Sixteen observers were recruited from the Munsell Color Science Laboratory community to evaluate the print-on-demand books. The age range was from 17 to 63, and all participants had normal color vision and acuity. Individual pages and books were evaluated in a controlled light booth (GTI EVS, D5000, 1800 lux). The viewing distance was not controlled, though observers viewed the individual pages on the table surface of the light booth and most observers evaluated the perfect-bound books while holding them in both hands, elbows bent. The experimental protocol was based on a Bartleson and Grum approach.[28]

3.2 *Experiment 2: Photobook Psychophysical testing*

3.2.1 *Photobook individual pages*

Pages from the seven photobook vendors were evaluated in a way analogous to the first experiment. The QS ranged from 1 to 7 (1=worst, 7=best).

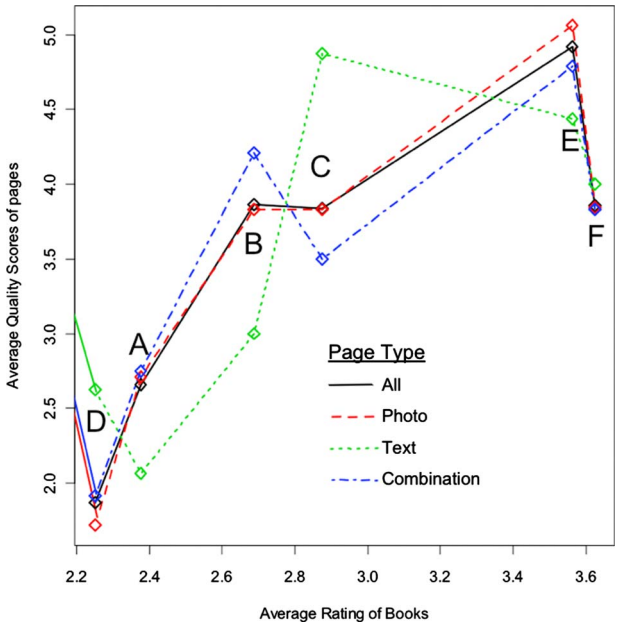


Fig. 5 Summary of average quality scores for all 10 pages (black solid line), six photo pages (red long dash), one text page (green short dash), and three combination pages (blue dot dash) shown on the vertical axis versus the ratings (on the horizontal axis) of the perfect bound books from the six vendors in experiment 1.

3.2.2 *Photobook bound books*

A similar process was used for the hardcover bound photobooks. However, the question about the price of lecture notes was changed to a question about the price of a recipe book to be sold as a fundraiser (called recipe price).

3.2.3 *Experimental details*

Nineteen observers participated in experiment 2. Eleven of these had previously been judges for the print-on-demand books. The additional eight participants were recruited from the broader Rochester Institute of Technology (RIT) community and were novice observers with respect to image-quality assessment. The age range was from 16 to 63, and all had normal color vision and acuity. Additional details were identical to the print-on-demand conditions.

4 Results and Analysis

4.1 *Experiment 1: Print-on-Demand Book Results*

4.1.1 *Print-on-demand psychophysics*

Figure 5 summarizes quality scores for the 10 pages and ratings of the perfect bound books from the six vendors. The horizontal axis shows the average ratings, starting from the right with the largest value of 3.63 for vendor F with a close second for vendor E. Both vendors E and F have average ratings that are statistically significantly (SS) higher than those of the remaining vendors. Vendor D (shown on the far left) received the lowest average rating of 2.25. The differences in average ratings among vendors A, B, C, and D are not SS except for the difference between C and D, based on the model discussed in Appendix A.

The vertical axis in Fig. 5 shows average QS of separate pages. The black line represents the overall average QS of all pages. Vendor E received the highest average QS of 4.9, which is SS higher than the next group of almost identical average QSs for vendors B, C, and F. Interestingly, vendors E and F both use HP Indigo printers, but their results vary. There are many potential reasons why two printers from the same manufacturer might vary in output. For example, the two Indigo printers are different models—E uses the 1000 series and F uses the 5000 series. Contrary to the subjective results, the Indigo 1000 specifications indicate lower dpi (800) and lpi (175) maximums than the 5000 series (812 and 230, respectively). However, software settings, including halftone parameters, calibration, machine status, and paper, can all impact quality responses. Both vendors B and C used iGen, printers and the results indicate these two vendors have similar image quality (based on QS and ratings). The remaining vendors (D and A) exhibit SS lower average QSs. Vendor D used an IBM DocuColor DC250. Participants commented that the pages from this book were grainy in appearance.

The red, green, and blue lines in Fig. 5 represent averages over the six photo pages, one text page, and three combination pages, respectively. The green line for the text page (with averaging only over observers) follows a pattern different from the other lines. With only one text page, it is difficult to make generalizations, but for this specific text page, vendor C has the highest text overall quality and vendor A has the lowest. The text page was designed to have matched sets of fonts in both 100% K and 50% K. Observation of the vendors' pages indicate that the density of the 50% K in vendor C's book was highest while the density for vendor A's 50% K was low and somewhat blurred. This was likely due to the fact that printer A mixed CMY inks for the 50% K specified. Observers equated darker printing of the 50% K text with higher quality.

The black, red, and blue lines in Fig. 5 (representing all pages, photo pages, and combination pages) follow very similar trends. This could indicate that overall quality response is overwhelmingly influenced by photo quality response. This question is worth further investigation. Also of importance is the suggestion that book content will influence which vendor may be the preferred option. For example, vendor C may have been the best option at the time of printing for books predominantly composted of text, whereas vendor E may have been the best for books with many images.

We can see from Fig. 5 that an increase in average book rating coincides with an increase in QS of single pages. One exception is the relationship between vendors E and F, where E has higher QS, but the two are practically equivalent in the book rating. The exception could be due to the difference in content between the perfect bound book and the individually removed pages. Regardless, the correlation coefficient of 0.83 indicates that the relationship between rating and QS is SS.

In order to investigate the prices that observers were willing to pay (vacation price and notes price), we calculated averages of both prices and ratings for each participant. We found that average ratings were not SS correlated with average prices, but the two types of price estimates were highly correlated with each other. The former makes sense for the magnitude estimation procedure we selected in which no price anchor was supplied to the observer. Thus, observers centered their price estimates around dif-

ferent mean prices. However, participants would presumably pay more for a better quality book. Hence, we used these prices as relative measures of book quality within the observers, but not among the observers. This in turn leads to a need to define normalized prices, which will be used in further analysis. The normalizing mechanism and its justification are explained in Appendix B.

After normalization, the ratings were highly SS correlated with the normalized vacation price (correlation coefficient equal to 0.95) and normalized notes price (correlation of 0.97). This indicates that observers associated range of image quality with range of price; perceptual differences in image quality influenced value assessment.

4.1.2 *Print-on-demand color gamut*

Two targets were incorporated into each book in order to assess the color gamut volumes. Neither was tagged or associated with ICC profiles. One target was a randomized version of the IT8.7/3 CMYK patches, while the second target consisted of RGB values divided into 10 levels such that the target contained 10-cubed (1000) patches. The inclusion of both CMYK and RGB encoding was to enable comparison of the printer image-processing paths. Printing RGB values requires the code values be converted into CMYK channels before printing. The purpose of including targets without embedded profiles (untagged images) was to probe the maximum color gamut volume the printer image pipeline generated for its given settings. Theoretically, targets without embedded profiles should represent maximum color gamut volume. In practice, many systems apply assumptions or profiles to untagged images. Printing systems can also use different rendering intents, i.e., perceptual versus relative colorimetric. Therefore, the printed targets may represent color gamut capability but can only be guaranteed to test the color gamut of the print result of the given book-printing job.

The spectral reflectance factors of the targets were measured with a GretagMacbeth Spectrolino SpectroScan. The spectra were then converted to CIELab values (D50, 2-deg standard observer) in order to obtain gamut volume numbers relevant to the viewing conditions used in the psychophysical experiment. Gamut volume was obtained using α shapes.[29]

Table 2 contains the estimated gamut volume for each of the vendors. The last column indicates how the CMYK volume compares to the RGB volume. Interestingly, the CMYK gamut volume for vendors B and D decreases relative to the RGB volume. These vendors may be treating CMYK code values as some type of color space rather than as straight CMYK or using different rendering intents (e.g., matching the original versus optimizing for the native CMYK). On inspection of the CMYK targets, it was apparent that the color ramps of CMYK were not pure ink laydown. This will fundamentally decrease the color gamut volume. The nearly identical gamut volume results, RGB to CMYK, for vendor A point to a simple direct conversion of RGB to CMYK or a gamut expansion method of the RGB data to extrapolate out to the CMYK gamut. The remaining vendors have the expected performance of a larger CMYK gamut volume.

Figure 6 contains two viewpoints of a plot of the CMYK and RGB color gamut volumes for vendor B that uses an iGen. The RGB volume (solid region) engulfs the CMYK volume (mesh). The other vendor that uses an iGen is vendor C. Figure 7 contains two viewpoints of the vendor C CMYK and RGB color gamut volumes. Note that vendor C's interpretation of the RGB target results in a smaller gamut volume than the CMYK target. The differences between the iGen gamuts from vendors B and C can occur because of ICC profile differences or selecting different rendering intents to use in the ICC profiles. One should also consider differences in paper (the properties of the paper itself or selection of an ICC profile for a given paper).

Table 2 Color gamut volume summary (in cubic CIELab units) for six print-on-demand vendors. RGB values are from an RGB-encoded target while CMYK values are from a randomized IT8.7/3 target. The last column indicates how the CMYK volume compares to the RGB volume.

Vendor designation	RGB	CMYK	CMYK volume % increase
A	307,433	309,056	0.53
B	429,179	313,516	–36.89
C	372,807	448,792	16.93
D	253,074	215,801	–17.27
E	401,549	448,567	10.48
F	224,045	273,149	17.98
Average:	331,348	334,814	

4.1.3 *Print-on-demand color accuracy*

Color accuracy was assessed using the simulated Macbeth ColorChecker target.[24] The target, made up of BabelColor sRGB code values, was embedded with an sRGB profile. Printed by each vendor, the target was measured as explained in Section 4.1.2. Table 3 contains the CIEDE2000 results for each vendor compared to the aim CIELab values calculated directly from the sRGB-encoded values in the target. The color difference values indicate that vendor B had the lowest color errors. See Fig. 8 for plots that show the color shifts from aim for vendor B. The vectors are very small for the majority of the 24 patches. The few colors with high color error do not seem to follow a trend. This could potentially be due to some system issues such as a bad value in a multidimensional lookup table in a profile.

Paradoxically, the two vendors with the highest overall image-quality results, vendors E and F, also had the highest color difference values. This reveals that accurate color does not necessarily equate to better image quality. Figure 9 illustrates that for vendor E all of the printed colors shift toward darker L* values. Thus, the higher color difference numbers are driven by what seems to be a more pleasing, darker (more saturated) version of the images.

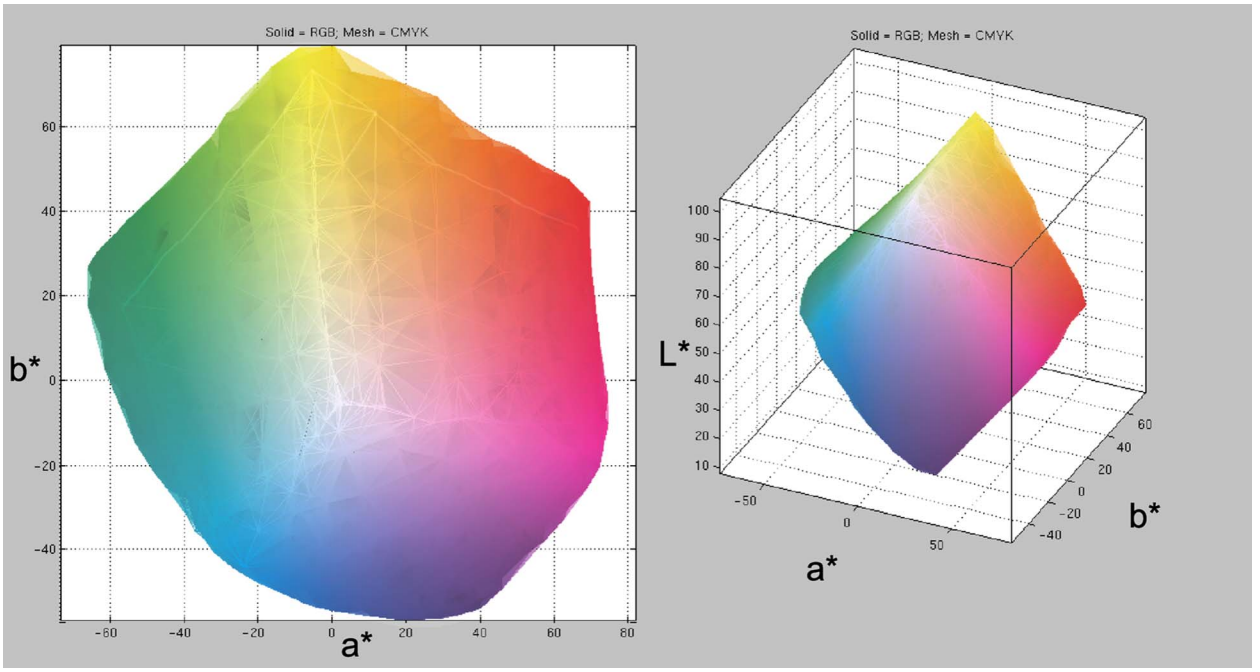


Fig. 6 Two viewpoints of vendor B color gamut volume (in cubic CIELab units). Solid is RGB gamut while engulfed mesh is CMYK gamut. RGB volume=429,179; CMYK volume=313,516.

4.1.4 *Print-on-demand profiling support*

Table 4 contains the evaluation of the ICC profiling interpretation for each vendor based on how each vendor rendered the target illustrated in Fig. 2. Vendors A and B failed the profiling probe, printing none of the four quadrants correctly. Inspection of the prints implies that the vendors are making an sRGB assumption regarding the RGB files passing through the image pipeline. Vendor D printed the version 4 profiles correctly, but not the version 2 profiles. Intuitively, this is surprising because the version 4 profile data structure choices are a superset of the version 2 profile data structure choices. Vendor E printed both version 2 profiles and one of the version 4 profiles correctly. It failed for a second version 4 profile in the sYCC quadrant. The sYCC

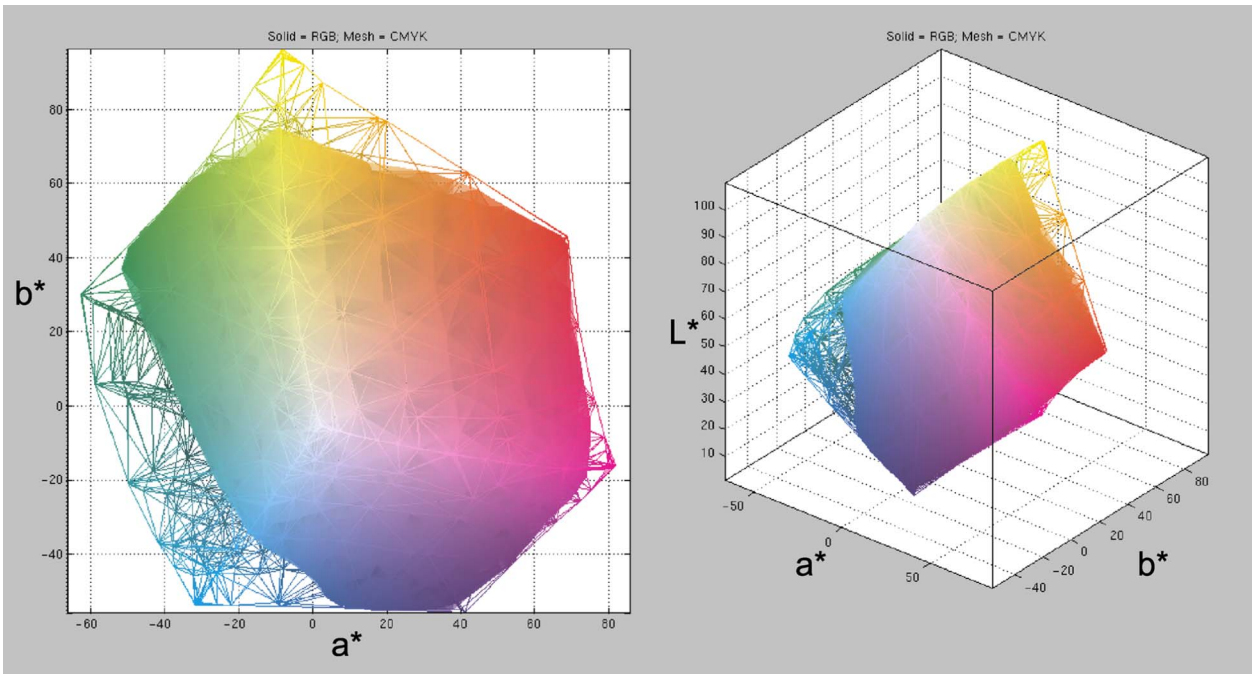


Fig. 7 Two viewpoints of vendor C color gamut volume (in cubic CIELab units). Solid is RGB gamut while mesh is CMYK gamut. RGB volume=372,807; CMYK volume=448,792.

Table 3 Simulated Macbeth ColorChecker CIEDE2000 Values for Print-on-Demand Books.

Vendor	Mean	Min	Max	σ
A	4.35	1.01	7.7	1.72
B	3.92	0.87	7.58	1.7
C	5.09	2.24	9.84	2.06
D	5.36	1.4	10.59	1.97
E	6.11	3.03	9.39	1.64
F	5.95	1.68	8.93	1.79

Table 4 Indication that the vendor used correct profile handling.

	ICC version 2		ICC version 4	
Vendor	Adobe RGB (1998)	GBR	e-sRGB	sYCC
A	No[a]	No	No	No
B	No[a]	No	No	No
C	Yes	Yes	Yes	Yes
D	No	No	Yes	Yes
E	Yes	Yes	Yes	No
F	Yes	Yes	Yes	Yes

[a]Difficult to assess.

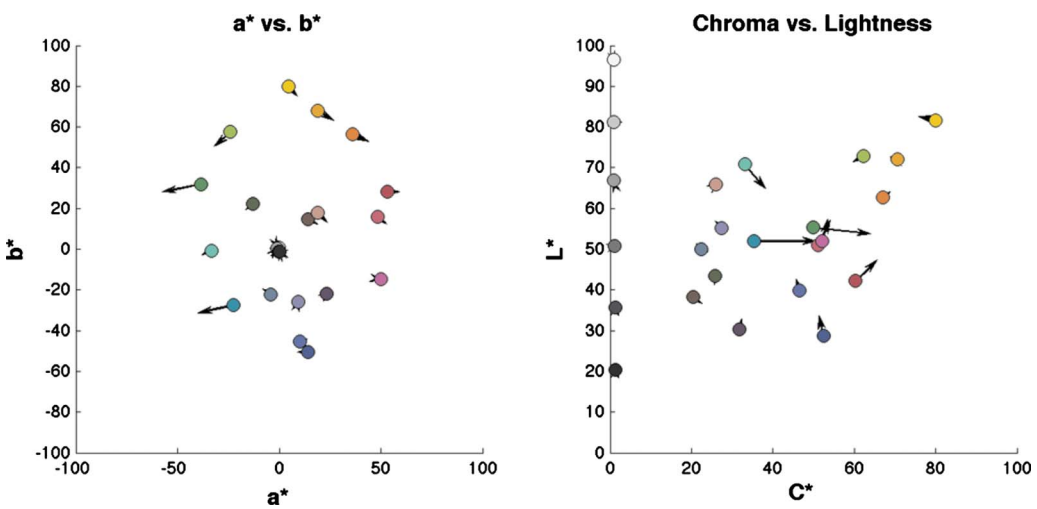


Fig. 8 Color accuracy shifts for vendor B's simulated Macbeth ColorChecker compared to aim sRGB values. CIEDE2000 mean=3.92, minimum=0.87, maximum=7.58.

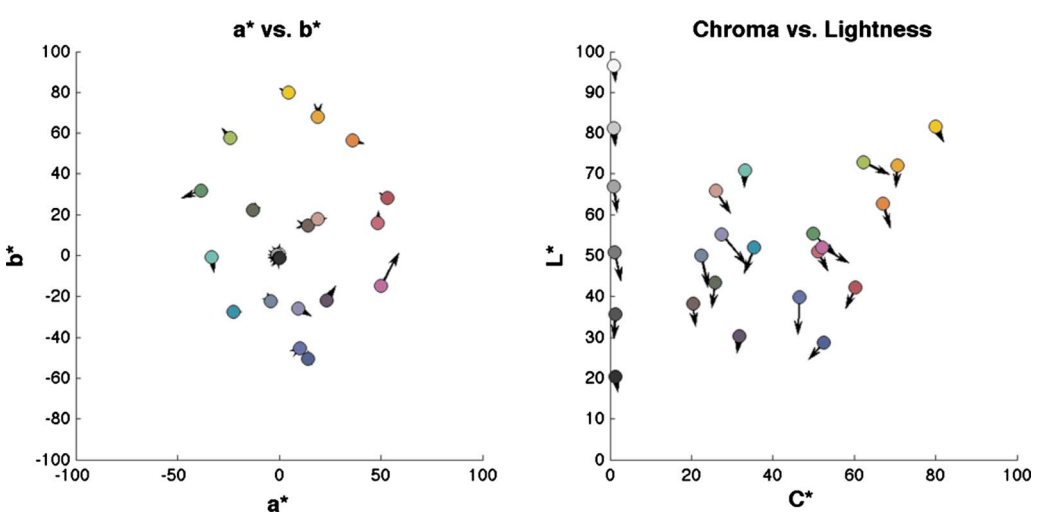


Fig. 9 Color accuracy shifts for Vendor E's simulated Macbeth ColorChecker compared to aim sRGB values. CIEDE2000 mean=6.11, minimum=3.03, maximum=9.39.

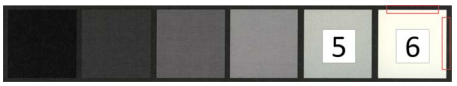


Fig. 10 Cropped scan of print-on-demand printed test target, showing sample edges chosen for edge MTF analysis of patch 6 in rectangle outlines.

profile utilizes the lutAtoBtype data structure, one that is not found in the version 2 specification. Thus, a processing package built to handle the version 2 data structures but that does not check the header to see if the profile is, in fact, version 2 will succeed in using the three other profiles but would not correctly handle the sYCC. Vendors C and F provided correct interpretation of all four embedded profiles. The various results for the six vendors indicate that ICC profile handling was still inconsistently handled in the one-off print-on-demand market at the time of our on-demand-book printing.

4.1.5 *Print-on-demand spatial analysis*

For each vendor, edges were chosen for obtaining measurements of an effective modulation transfer function (MTF). The MTF is a well-established measure of the transfer of image information or detail during image capture, display, or printing.[30] MTF is often used in computed measures of image sharpness for text, graphics, and rendered image content. Various methods have been used to measure the MTF of printing systems, including those based on image noise, periodic signals such as sine waves,[31,32] lines, and edges.[32,33] For the studies in this paper, we chose to use the method based on the analysis of printed edges. This method, often called edge-gradient analysis, has the advantage of using a simple, compact edge feature. In addition, a widely used implementation, based on slanted edges,[33,34] can be used, or easily adapted for edges that are vertical or horizontal on the page, rather than rotated from these orientations. For the case of edge-based printer MTF measurement, our input target feature is a noise-free edge in the input digital image. Our corresponding measured output includes the influence of imaging material (toner or ink), paper, and rendering of the edge.

Figure 10 highlights an example of the horizontal and vertical edges selected within the neutral patches of the printed simulated Macbeth ColorChecker (which has been rotated 90 deg here from the printed direction). Analysis was performed on horizontal and vertical edges of patches 5 and 6. For this method, adapted from that developed for scanner evaluations,[34] the edge profile normal to the edge is obtained by averaging along the edge. The first derivative of the sampled profile is then computed and taken as the effective point-spread function. The discrete Fourier transform of this vector is computed, and after scaling to unity at zero frequency, this is the measured MTF.

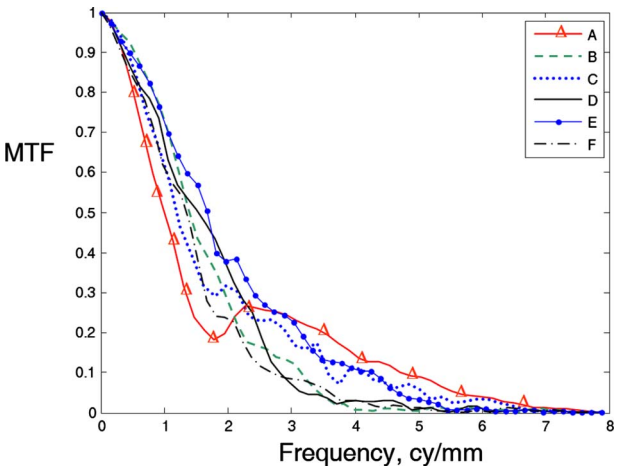


Fig. 11 Edge-based printer MTF for horizontal edges of patch 5.

Figure 11 contains example results of edge-MTF data from patch 5 horizontal edges for all vendors. We should note that two aspects of the printed edge features can influence these results. The first is the shape and variation in location of colorant on the print. This could include toner clumping in electrophotography. The second effect can be due to the spatial rendering, or digital halftoning, that results in variations normal to the edge.

A measure of image sharpness, cascaded modulation transfer (CMT) acutance,[35] was computed from these MTF results. Details on calculating this metric appear in Appendix C. The CMT results are shown in Table 5. Vendor E is

Table 5 CMT Acutance values for print-on-demand books from neutral patches of the simulated Macbeth ColorChecker target. V=vertical direction, H=horizontal direction.

Vendor	Patch 5 V	Patch 5 H	Patch 6 V	Patch 6 H	Overall average
A	81.65	78.68	84.12	80.68	81.28
B	82.17	84.06	77.33	83.69	81.81
C	83.86	82.16	88.72	77.13	82.97
D	81.55	83.18	84.31	83.04	83.02
E	86.66	86.01	87.04	85.58	86.32
F	86.25	80.98	84.44	81.11	83.20

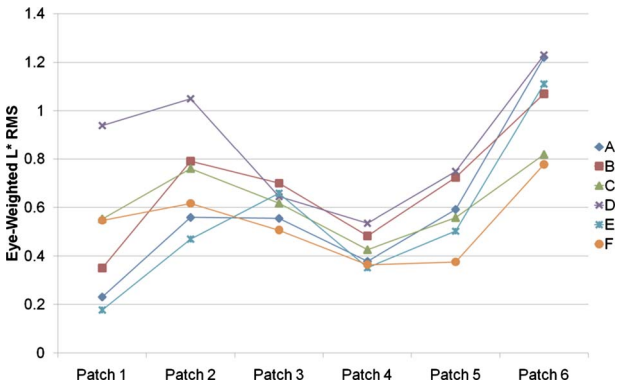


Fig. 12 Eye-weighted rms L* values of patches for each print-on-demand vendor. Higher values model higher perceived noise.

most consistent in acutance and has the highest overall average. This vendor performed well for both rating and QS.

In order to assess noise, the scanned data for all the patches shown in Fig. 10 were transformed to CIELab coordinates (2 deg, D50). An eye-weighted root mean square (rms) L* (called W-rms for briefness) was calculated to obtain a noise metric related to perception. Details on the calculations are in Appendix D.

The noise results are shown in Fig. 12. The patch-to-patch trends of the noise for each vendor are very similar, an indication that noise is density-specific for printed neutrals. Vendor D exhibits the highest W-rms for most patches. Many participants commented about the grainy appearance of the printed pages for vendor D, and the rating and QS values are low. Vendor F has low W-rms, indicating the patches have low noise levels. This vendor has high rating and QS values.

4.1.6 *Relationships between the objective and subjective assessments of quality*

In Section 4.1.1, we discussed some subjective measures of book quality, such as QS, rating, normalized vacation price, and normalized notes price. Now we want to investigate to what extent those subjective measures are impacted by objective quality measures discussed in Sections 4.1.2–4.1.5. For each of the objective measures, we calculated the average value for each vendor and then the correlation coefficients (based on the resulting six observations) with the average values of the four subjective measures. We found the highest correlations (shown in Table 6) when using the W-rms objective quality measure discussed in Section 4.1.5.

Table 6 Correlation coefficients between the W-rms average values and averages of subjective measures of book quality.

	Subjective measures of book quality			
Objective measure	Quality score	Rating	Normalized vacation price	Normalized notes price
W-rms	−0.73	−0.79	−0.75	−0.68

Table 7 Estimated true correlation coefficients between the eye-weighted rms L^* average values and subjective measures of book quality.

	Subjective measures of book quality			
Objective measure	Quality score	Rating	Normalized vacation price	Normalized notes price
W-rms	−0.81	−0.95	−0.89	−0.78

Not all correlations in Table 6 are SS. For only six observations, the smallest correlation significant at the level $\alpha=0.05$ is 0.82.[36] However, these calculations do not take into account the fact that each of the six vendors' scores is an average of several observations. In order to deal with this issue, we used linear mixed-effects models (see Appendix E for details) and estimated the true correlations as shown in Table 7.

The statistical significance of the correlations in Table 7 was assessed based on the likelihood ratio test (see Appendix E). The correlations are significant at level $\alpha=0.05$, except for that of W-rms and normalized notes price (value of −0.78), which would be significant only at $\alpha=0.075$.

These results suggest significant correlations between the objective and subjective measures of quality within experiment 1. However, due to the small number of vendors investigated, the statistical significance of the results is not very high. It also means that in hypothetical replications of this experiment, the correlations may fall and not be SS (in statistics, it is described as a low power of testing).

Similar analysis of the relationship between the objective metrics and psychophysical results indicate that correlation between CMT acutance and the four subjective measures of book quality is SS. However, the objective metrics of color gamut and color accuracy are not correlated with the subjective metrics. This may indicate that color gamut and color accuracy are not optimal metrics for predicting perceived image quality.

4.2 *Experiment 2: Photobooks Results*

4.2.1 *Photobook psychophysics*

Figure 13 is organized in a way similar to Fig. 5. That is, it summarizes QS for the 10 pages and ratings of the hardcover bound books from the seven vendors in experiment 2. The horizontal axis shows the average ratings, starting from the right with the largest value of 4.11 for vendor I being SS larger than the remaining vendors. The group of the remaining vendors does not show any SS differences except for the SS difference between K and L, based on the model discussed in Appendix A.

The green line for the text page again follows a pattern different from the other lines. As in experiment 1, it is difficult to make generalizations from only one text page. For this specific text page, the text was designed to have both 100% K and 50% K text. Observation of the vendors' pages indicates that the density of the 50% K in vendor K's book is highest and nearly indistinguishable from the 100%

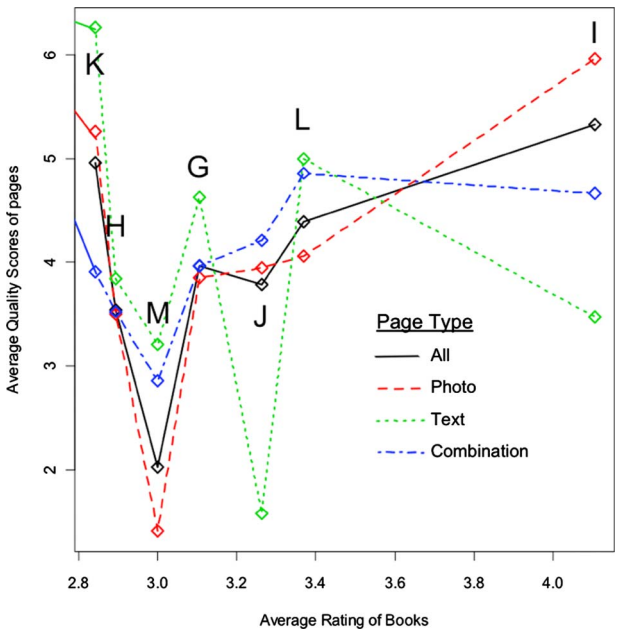


Fig. 13 A summary of average QS for all 10 pages (black solid line), six photo pages (red long dash), one text page (green short dash), and three combination pages (blue dot dash) shown on the vertical axis versus the ratings (on the horizontal axis) of the hardcover bound books from the seven vendors in experiment 2.

K. In addition, the font is very sharp. The reproduction of the font on vendor J's text page has significantly low acuity compared to other photobook vendors.

The black, red, and blue lines in Fig. 13 (representing all pages, photopages, and combination pages, respectively) are not as close to each other as they were in experiment 1. We can see from Fig. 13 that an increase in average book rating does not necessarily coincide with an increase in QS of single pages, and the correlation between these two subjective quality measures is not SS (correlation coefficient =0.51). Several factors could be contributing to this. The most probable reason is that each page for each vendor was assembled with the vendor software, resulting in sets of pages that were not always identical in layout (vendor page templates differed). Whereas in experiment 1, the same page layouts were uploaded as PDF files to each vendor, the individual files were hand-placed in page templates for experiment 2. This situation in experiment 2 also allowed for vendor-to-vendor differences in applying image processing to the pages and images. The appearance of pages of some vendors pointed to the possibility of localized image processing algorithms.

Vendor M exhibits low QS relative to rating, while vendor K exhibits high QS. This points to the difference in assessing individual pages versus complete hardcover bound books. More specifically, the individual pages in vendor M tended to highlight the image-quality degradations due to an overall purple-blue shift from this vendor (see Section 4.2.3 for more details). Vendor K exhibited nonuniformity in flat fields. However, this image-quality concern was more noticeable in the bound book content (used for rating) versus the individual pages (used for QS).

As in the experiment 1 context, both types of estimated prices were normalized. Again, the ratings were highly SS correlated with the normalized vacation price (correlation of 0.98) and normalized recipe book price (correlation of 0.98).

Table 8 Color gamut volume summary (in cubic CIELab units) for seven photobook vendors.

Vendor designation	RGB
G[a]	384,064
H	307,032
I	379,316
J	342,189
K[a]	391,134
L[a]	343,332
M[a]	323,980
Average:	353,007

[a]RGB target missing patches. See body text for more details.

4.2.2 *Photobook color gamut*

Because many of the software interfaces for the photobooks were not set up to accept CMYK files, color gamut was assessed with only the untagged RGB target described in Section 4.1.2. The process to obtain the gamut volumes in cubic CIELab units was the same as the experiment 1 analysis. Table 8 contains the results for each photobook vendor. Because of the constraints of print rendering and placing the RGB target into several vendor-page templates that were "fill" only, target patches were digitally cropped or physically lost in the binding glue in four of the seven vendors. Specifically vendors G and M were missing 25 patches, vendor L was missing 75 patches, and vendor K was missing 200 patches. Therefore, the actual gamut volume is most likely larger than what is represented in the table for these vendors since the complete RGB target contains 1000 patches.

The results in Table 8 indicate the color gamut of the print result for each given book-printing job. Ideally, a metric of the maximum color gamut capability was desired. However, this level of control was not available. The largest reported color gamut is for vendor K, the vendor with the lowest rating. The next-lowest rating is for vendor H, which has the smallest reported color gamut. These results indicate that color gamut volume may not be an optimal test for predicting perceived image quality. On average, the overall color gamut of the photobooks is ~6% larger than that of the print-on-demand books (see Table 2). This overall difference is most likely even greater due to the issue of not obtaining all 1000 target patches for all vendors in experiment 2. However, print-on-demand vendors B and E do have larger color gamuts (429,179 and 401,549, respectively) than vendor K with the largest photobook gamut (391,134).

Table 9 Simulated Macbeth ColorChecker CIEDE2000 values for Photobooks.

Vendor	Mean	Min	Max	σ
G	4.56	1.57	7.61	1.78
H	5.33	0.94	12.48	2.56
I	3.99	0.79	7.72	2.08
J	4.92	1.43	11.22	2.24
K	5.5	2.09	11.18	2.61
L	4.78	0.96	12.59	2.36
M	8.89	4.56	13.25	2.55

4.2.3 *Photobook color accuracy*

Color accuracy was assessed using the simulated Macbeth ColorChecker target as in Section 4.1.3. Table 9 contains the CIEDE2000 results for each vendor compared to the aim CIELab values calculated from the sRGB-encoded values in the target. The color difference values indicate that vendor M has the largest color accuracy error (8.89), where even the minimum CIEDE2000 is 4.56.

Studying the vendor M color shifts in Fig. 14 more closely confirms the overall purple-blue hue shift that was observed by many participants. The neutrals exhibit a strong increase in chroma. The behavior of other vendors was not as acute, and ranges were similar to the print-on-demand results for experiment 1.

Table 10 Indication that the vendor used correct profile handling. Images with e-sRGB and sYCC profiles were not accepted by vendor I.

	ICC version 2		ICC version 4	
Vendor	Adobe RGB (1998)	GBR	e-sRGB	sYCC
G	Yes	Yes	Yes	Yes
H	Yes	Yes	Yes	Yes
I	Yes	Yes	NA	NA
J	No	No	No	No
K	No	No	No	No
L	No	No	No	No
M	Yes	Yes	Yes	Yes

4.2.4 *Photobook profiling support*

Similar to the state for the print-on-demand books, not all photobook vendors handle ICC profiles correctly (see Table 10). Vendors J, K, and L did not use correct profile handling, while vendors G, H, and M did. For vendor I, the photobook software would not accept the images with e-sRGB or sYCC profiles. This might imply that the vendor was aware that these are not ICC version 2 profiles and chose not to address these profiles. Vendor I correctly interpreted only the ICC version 2 profiles.

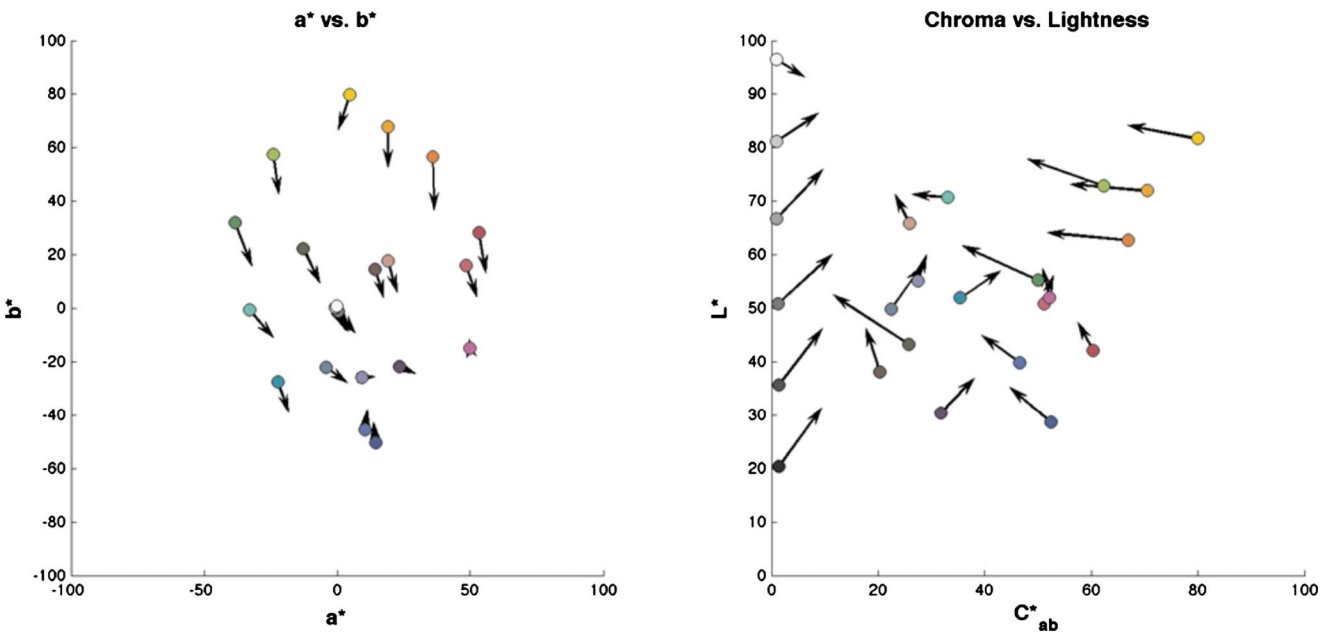


Fig. 14 Color accuracy shifts for vendor M's simulated Macbeth ColorChecker compared to aim sRGB values. CIEDE2000 mean=8.89, minimum=4.56, maximum=13.25.

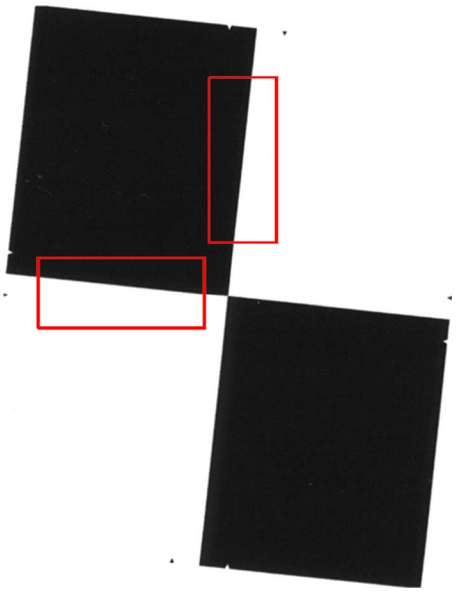


Fig. 15 Printed elements used to evaluate the photobook printer MTF using the slanted-edge method. The rectangle outlines indicate the analysis regions.

4.2.5 *Photobook spatial analysis*

The printer edge MTF measurement for the photobooks was made, as before, using edges. However, in this experiment the edges were slanted from the print direction as shown in a scan of the printed graphic elements in Fig. 15. This slanted edge was chosen for experiment 2 as a more accurate assessment of the sharpness than the target used in experiment 1. The analysis regions for the vertical and horizontal measurements are shown by the red rectangles. For this method, the edge profile normal to the edge is obtained by averaging along the edge. The measured MTF of these edges as well as CMT acutance were obtained in a similar way as with the print-on-demand books.

Table 11 contains the acutance results for the photobook vendors. These values are all higher than those in experiment 1. However, the difference is most likely due to the difference in targets rather than absolute sharpness differences. Vendor L exhibits the lowest average acutance, while vendor M exhibits the highest. These results are not intuitive based on the psychophysical data. However, the slanted edge was placed in each photo template and control over resizing was left up to the vendors. This may have altered native system performance relative to sharpness measurements.

Table 11 CMT Acutance values for photobooks from neutral patches of the simulated Macbeth ColorChecker target.

Vendor	Vertical	Horizontal	Overall average
G	88.77	92.79	90.78
H	89.73	90.11	89.92
I	90.22	91.77	91.00
J	91.89	90.96	91.43
K	91.55	91.71	91.63
L	89.74	89.85	89.80
M	91.44	93.37	92.41

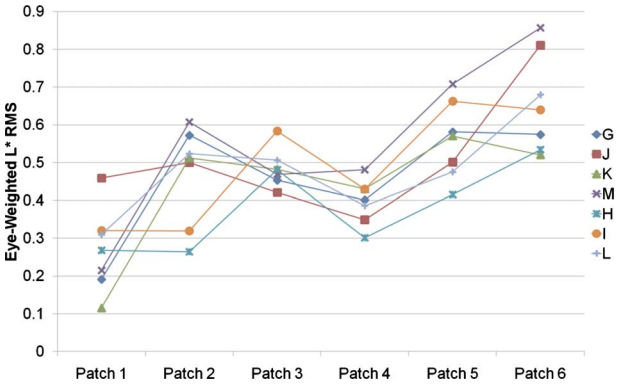


Fig. 16 Eye-weighted rms L⋆ values of patches for each photobook vendor.

Figure 16 contains the W-rms results for the photobook vendors, an indication of the density-dependent behavior of noise. As in experiment 1, the calculations were made from the scanned neutral patches of the Macbeth ColorChecker chart. Although most vendors follow similar noise trends per patch, the results are more variable than those of the print-on-demand books (see Fig. 12). However, the photobook results are lower overall, indicating that the print-on-demand patches were noisier overall. Vendor M has high W-rms values (high noise). This poorer behavior is also seen in the ratings and QS. Interestingly, vendor I exhibits higher W-rms levels in middle and upper patches. As shown in Fig. 13, vendor I is SS higher in ratings and QS than other vendors—a low correlation with the noise results.

4.2.6 *Relationships between the objective and subjective assessments of quality*

In this section, we use the methods discussed in Section 4.1.6 to assess the relationships between the objective and subjective results. The analysis indicated that none of the objective metrics were SS correlated with the psychophysical results. Although this may raise questions regarding the results and analysis, several factors contributed to variability in both objective and subjective metrics. Because the photobooks were assembled with vendor software and templates, precise control over layout and target sizing were limited and thus differed for each photobook. For example, the layouts of the multiphoto pages were not identical in placement or size between different vendors. Some targets appeared to be resized prior to printing, changing the native resolution and intent of spatial evaluation. Some targets appeared to have localized image-processing algorithms applied that would cause unpredictable effects on performance, such as color accuracy. Thus, photobook differences in the page and book-based psychophysical testing would have been accentuated compared to the print-on-demand

books, which were printed from identical PDF content. The printer manufacturers and models were unknown for experiment 2. In experiment 1, there were two pairs of very similar printers that may have led to SS correlation with the spatial metrics and subjective results, which may not have been present in this second experiment. From the statistical point of view, we would expect large variability from experiment to experiment (resulting in both significant an insignificant results) when small numbers of vendors are used (six and seven in this study).

5 Conclusions

Six one-off print-on-demand vendors and seven photobook vendors were assessed for overall image quality. Although the photobooks were assessed with higher overall image quality, their performance was also more variable than the print-on-demand books. In order to deal with small samples and adjust for some sources of variability, we proposed here an original approach using latent mixed-effect models. This approach allowed us to compare the objective and subjective metrics.

Several elements were ascertained from the psychophysical results. Rank-order results indicate that photographic image quality on a page may overwhelm the quality of text when assessing the overall quality of a book. Because some publishers had different quality results for photos versus text, book content may be the deciding factor for selection of one-off vendors of choice. We proposed a new approach to normalizing price estimates so that they could be used in further analysis in a meaningful way. Overall image-quality ratings and estimated price point were highly correlated, indicating that observers associated range of image quality with range of price; perceptual differences in image quality influenced value assessment.

Three color metrics were used to evaluate the books: color gamut, color accuracy, and accuracy of ICC profile usage. Analysis of the color gamut and color accuracy indicated that high overall image quality could be obtained from measurably different color performance. Color metrics were not SS correlated with subjective results and point to the challenge of determining color metrics, which predict overall image quality. Also, the same printer hardware among different vendors did not guarantee similar image quality; software and settings appear to have significant impact on image quality. In fact, color gamut comparisons of the print-on-demand books revealed that untagged CMYK gamuts are sometimes smaller than untagged RGB gamuts, which points to specific assumptions built into the vendors' software that can influence quality. Finally, use of a target to test accuracy of ICC profile usage revealed that ICC profile handling was still inconsistently handled in the one-off print-on-demand and photobook markets at the time of printing; expansion beyond sRGB assumptions was still in process.

Spatial evaluations were made to assess noise (eye-weighted rms L*) and sharpness (CMT acutance). The correlation of these objective metrics with the psychophysical results was SS for experiment 1 in which all seven vendors printed a PDF with identical page layouts. In experiment 2, the correlation of the objective metrics and psychophysical results were not SS. Both the objective metrics and subjective results were more variable in experiment 2. Causes for the higher variability are due, in part, to the book-to-book differences inherent in using seven different book-assembly software packages, image processing (which may have been localized for some vendors), and a potentially greater number of printer manufacturers and models.

Acknowledgments

G. J. Woolfe wrote the fundamental aspects of the Matlab code to calculate color gamut volume. Mark V. Giglio provided test target measurement support. Portions of this research were funded by Eastman Kodak Company and Thomas Lianza.

Appendix A. Statistical Models for Investigation of Factors Impacting Quality Assessment

Let us start with a discussion of ratings describing quality of books. Each rating depends on the book being evaluated (so indirectly on the vendor), on the observer evaluating the book, and on some other unobserved factors. This can be expressed by the following statistical model:

$$Y_{ij} = \beta_i + \tau_j + \varepsilon_{ij},$$

where Y_{ij} is the rating given for the i'th vendor's book by the j'th observer, β_i is the average rating of the i'th vendor's book, τ_j expresses variability among participants, that is, how different the j'th observer is from an average observer. The "error" term ε_{ij} accounts for all the remaining factors, which are not observed. In the above model, there are two factors—vendor and observer—and β_i, τ_j are their main effects. The main effect τ_j is assumed to be random, and the model is called a mixed-effects model.

The QS analyzed in Sections 4.1.1 and 4.2.1 depend on three factors: vendor, observer, and page being evaluated. Because each observer in experiment 1 assigns QS 1–6 to the vendors' books, the average of those scores is always 3.5. Hence, we are not able to estimate the main effects of the observer and the page. We are then left with a simple one-way analysis of variance model

$$Y_{ij} = \beta_i + \varepsilon_{ij},$$

where Y_{ij} is the QS given for the i'th vendor's book by the j'th observer and β_i and ε_{ij} are the same as explained before.

In both of the above models, the significance of differences among vendors is assessed based on multiple comparisons of the main effects β_i.

Appendix B. Normalization of Prices

Here we explain the normalization of prices as used in Sections 4.1.1 and 4.2.1. The same process is followed for both types of prices in both experiments. Let X_{ij} be the price given for the i'th vendor's book ($i=1,\ldots,6$) by the j'th observer ($j=1,\ldots,16$). We calculated the averages and standard deviations over books for all observers, that is, $\bar{X}_{.j}$ and s_j, respectively. Figure 17 shows the scatter plots of those standard deviations versus averages for both types of prices and both experiments. We can clearly see a pattern, mostly linear. When a simple linear regression model is fitted to those data, the fitted value for the standard deviation can be interpreted as the average variability for a given

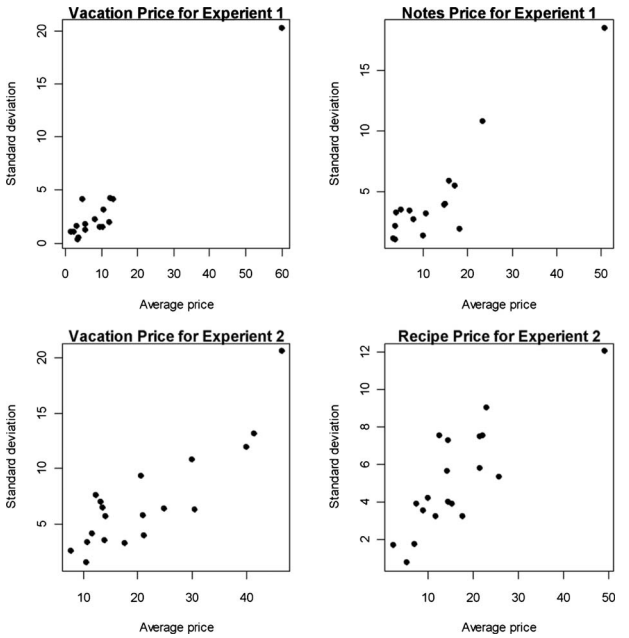


Fig. 17 Scatter plots of standard deviations versus averages (calculated over the number of vendors) for all observers (one dot represents one observer) for both types of prices and both experiments.

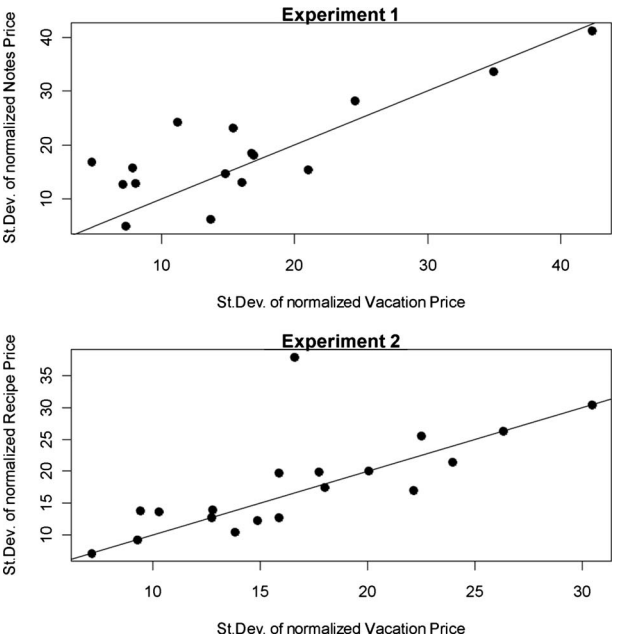


Fig. 18 Scatter plot of standard deviations of the normalized notes price versus those of the normalized vacation price (calculated over the number of vendors) for all observers. Each dot represents one observer. The straight line shows where the two standard deviations are equal to each other.

mean value. The observers represented by points above the fitted line can be regarding as having opinions stronger than those of an average observer. On the other hand, points below the fitted line represent participants who do not feel strongly about the differences in quality of the vendors' books. Consequently, we want to normalize the prices, so that each observer has the same average, say M, but possibly different standard deviations reflecting the strength of their options. It does not matter for the analyses performed in this paper what the value of M is; it could be \$50, for example. The normalized value is calculated from the formula

$$\frac{X_{ij}}{\bar{X}_{.j}}M,$$

which means that each price given by the j'th observer is divided by the observer's average and multiplied by M. In order to see how this formula works, consider a linear fit to the data going through the origin. If a observer is represented by a point on that fitted line, then after the normalization, the point will be moved to another point on that line with the horizontal coordinate value of M. That way, the relative position of that observer will be kept in the middle range of the strength of options (as expressed by the standard deviation). In a similar way, the relative position of any participant in relation to other observers will not change. Figure 18 shows the relationship between standard deviations of the normalized notes price versus that of the normalized vacation price ($M=50$ was used). Each dot in the plot represents one observer. The straight line shows where the two standard deviations are equal to each other. There is a slight tendency for observers to be more opinionated about the notes price versus the vacation price (with some points further above the line). The two standard deviations are highly correlated ($r=0.82$), which further supports our approach.

Appendix C. Method for Calculating CMT Acutance

A feature of sharpness, CMT acutance,[35] was computed from the MTF results of the book vendors as follows. Each MTF, which represents the spatial frequency response (SFR) of the system, was multiplied frequency-by-frequency by a visual contrast sensitivity function (CSF). The CSF used is that due to Daly[37] for a viewing distance of 33 cm. This weighted vector was then integrated to a single number representing the area under the weighted curve

$$a = \int_0^{f\,\max} \mathrm{SFR}_{\mathrm{system}}(f)\mathrm{CSF}(f)df.$$

This value was then scaled by the corresponding value for an ideal system, where $\mathrm{SFR}_{\mathrm{system}}=1$. The visual response value, R, is the ratio

$$R = \frac{a}{\int_0^{f\,\max} \mathrm{CSF}(f)df}.$$

This value was then modified to give the computed sharpness value for the system,

$$\mathrm{CMT} = 100 + 66\log_{10}(R),$$

where R is constrained to the range $[0,1]$. The CMT results for the print-on-demand books are shown in Table 5, while

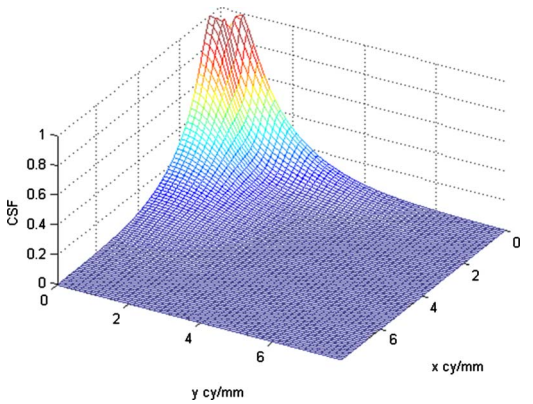


Fig. 19 Contrast sensitivity function used for visually weighted noise evaluation.

the results for the photobooks are shown in Table 11.

Appendix D. Method for Calculating Eye-Weighted rms L*

An eye-weighted rms L* (called W-rms for briefness) was calculated to obtain a noise metric related to perception. The two-dimensional noise-power spectrum (NPS) was computed from the L* image array on uniform regions of 350×350 pixels. This NPS array was weighted frequency by frequency by a visual CSF. For a random variable, the variance statistic, σ^2, is equal to the integral of the corresponding NPS. This property was used to compute the variance of the L* array as the sum of the NPS

$$\sigma^2 = \Delta f_x \Delta f_y \sum_{i=\min 1}^{\max i} \sum_{j=\min j}^{\max j} \mathrm{CSF}_{i,j}^2 \mathrm{NPS}_{i,j},$$

where Δf_x and Δf_y are the spatial frequency sampling intervals for the measured NPS, and the summations are over the range of frequencies in the x and y directions. The square root of the above variance measurement was taken as the eye-weighted rms L* value. The CSF used is shown in Fig. 19.

Appendix E. Statistical Models for Investigation of the true Underlying Correlation

Here we explain the statistical methodology behind the estimation of the true correlation coefficients between the W-rms objective quality measure and the subjective measures (as used in Sections 4.1.6 and 4.2.6). As an example, we will use rating as a subjective measure. Recall the mixed-effects model used in Appendix B

$$Y_{ij} = \beta_i + \tau_j + \varepsilon_{ij}.$$

In a similar fashion, we can define a mixed-effects model for W-rms

$$X_{ij} = \beta_i^* + \gamma_j + \varepsilon_{ij}^*,$$

where X_{ij} is the W-rms calculated for the j'th patch sample ($j=1,\ldots,4$ for horizontal and vertical samples from two patches) from the i'th vendor's book ($i=1,\ldots,6$), β_i^* is the average rating of the i'th vendor's book, γ_j expresses variability among patch samples. As before, the "error" term ε_{ij}^* accounts for all the remaining factors, which are not observed.

This time, we need to treat the effects β_i and β_i^* as random and assume a general covariance matrix for the random vector $[\beta_i, \beta_i^*]^T$. Note that β_i's are the latent (unobserved) effects of the vendor on the ratings, and β_i^*'s are the latent (unobserved) effects of the vendor on the W-rms. Hence, the correlation between β_i and β_i^* is the true correlation between ratings and W-rms over the population of vendors. The same methodology can be used when ratings are replaced by QS, normalized vacation prices, or normalized notes prices, except that the effects τ_j can be removed in these cases.

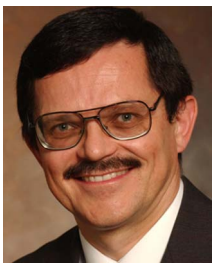
More information about the models used here can be found in Ref. 38. Calculations were performed in the statistical language R using the function lmer with the restricted maximum likelihood estimation.

References

1. A. Chowdry, "Digital color printing and the NexPress approach," in *Proc. of NIP17*, p. 267, IS&T, Springfield, VA (2001).
2. J. Simal, "On-demand printing transforming museum visitors experience," *Proc. of DPP2005*, p. 141, IS&T, Springfield, VA (2005).
3. A. C. Hübler, "Digital high volume printing—breakthrough for print-on-demand?," in *Proc. of NIP*, Vol. 15, p. 1, IS&T, Springfield, VA (1999).
4. C. Lahanier, A. de Polo, A. Minodier, and J. Misselis, "Global art on demand initiative," in *Proc. of Photofinishing Technologies*, p. 98, IS&T, Springfield, VA (2004).
5. Lulu, ⟨http://www.lulu.com⟩ (last accessed Nov. 2007).
6. UBuildABook, ⟨http://www.ubuildabook.com⟩ (last accessed Dec. 2007).
7. Flexpress, ⟨http://www.flexpress.com⟩⟩ (last accessed Nov. 2007).
8. NetPublications, Inc., ⟨http://www.netpub.net⟩ (last accessed Nov. 2007).
9. StarNet Digital Publishing, ⟨http://www.starnetdp.com⟩ (last accessed Nov. 2007).
10. R and R Images, ⟨http://www.randrimages.com⟩ (last accessed Nov. 2007).
11. Aardvark Global Publishing, ⟨http://www.aardvarkglobalpublishing.com⟩ (last accessed Mar. 2008).
12. Snapfish by HP, ⟨http://www.snapfish.com⟩ (last accessed Apr. 2008).
13. Shutterfly, ⟨http://www.shutterfly.com⟩ (last accessed Apr. 2008).
14. Blurb, ⟨http://www.blurb.com⟩ (last accessed Apr. 2008).
15. Apple iPhoto, ⟨http://www.apple.com/ilife/iphoto⟩ (last accessed Apr. 2008).
16. Apple Aperture, ⟨http://www.apple.com/aperture⟩ (last accessed Jan. 2008).
17. Kodak Gallery, ⟨http://www.kodakgallery.com⟩ (last accessed Dec. 2007).
18. Picaboo, ⟨http://www.picaboo.com/⟩ (last accessed Dec. 2007).
19. P. Sandhaus, S. Thieme, and S. Boll, "Processes of photo book production," *Multimedia Syst.* **14**, 351–357 (2008).
20. N. Henze and S. Boll, "Snap and share your photobooks," in *Proc. of 16th ACM Int. Conf. on Multimedia*, Vancouver, Oct. 26–31, 2008, MM '08, ACM, New York, pp. 409–418 (2008).
21. S. Boll, P. Sandhaus, A. Scherp, and S. Thieme, "Metaxa—context- and content-driven metadata enhancement for personal photo books," in *Proc. of 13th Int. Multi-Media Modeling Conf.* (MMM), Singapore, Springer, New York (2007).
22. E. Fredericks, P. Bajorski, and M. R. Rosen, "Preferred color correction for mixed taking-illuminant placement and cropping," in *Proc. of CIC*, p. 198, IS&T, Springfield, VA (2009).
23. ANSI, "Graphic technology—Input data for characterization of 4—color process printing," ANSI IT8.7/3 (1993).
24. D. Pascale, RGB Coordinates of the Macbeth ColorChecker, ⟨http://www.babelcolor.com/download/RGB%20Coordinates%20of%20the%20Macbeth%20ColorChecker.pdf⟩ (2006).
25. ISO, ISO 12640-3:2007, Graphic technology—prepress digital data exchange, Part 3: CIELAB standard colour image data (CIELAB/SCID) (2007).
26. Adobe Systems, ⟨http://www.color.org/version4html.xalter⟩ (2003).
27. L. Borg, ⟨http://www.color.org/version4html.xalter⟩ (2003).
28. J. Bartleson and F. Grum, eds., *Optical Radiation Measurements*, Vol. 5, p. 462, Academic Press, New York (1984).
29. T. J. Cholewo and S. Love, "Gamut boundary determination using

alpha-shapes," in *Proc. of CIC*, p. 200, IS&T, Springfield, VA (1999).
30. C. J. Dainty and R. Shaw, chapter 5 in *Image Science*, Academic Press, New York (1974).
31. W. Jang and J. P. Allebach, "Characterization of printer MTF," *Proc. SPIE* **6059**, 1–12 (2006).
32. A. J. Lindner, N. Bonnier, C. Leynadier, and F. Schmitt, "Measurement of Printer MTFs," *Proc. SPIE* **7242**, 72420N (2009).
33. P. D. Burns, "Slanted-edge MTF for digital camera and scanner analysis," *Proc. of PICS Conf.*, IS&T, p. 135 (2000).
34. ISO, ISO 16067-1:2003, Photography—Spatial resolution measurements of electronic scanners for photographic images, Part 1: Scanners for reflective media (2003).
35. R. G. Gendron, "An improved objective measure for rating picture sharpness: CMT acutance," *J. Soc. Motion Pict. Telev. Eng.* **82**, 1009–1012 (1973).
36. J. L. Devore, *Probability and Statistics for Engineering and the Sciences*, p. 544, Brooks/Cole, Pacific Grove, CA (2004).
37. S. Daly, "The visible differences predictor: an algorithm for assessment of image fidelity," Chapter 13 in *Digital Images and Human Vision*, A. B. Watson, Ed., p. 149, MIT Press, Cambridge, MA (1993).
38. J. C. Pinheiro and D. M. Bates, *Mixed-Effects Models in S and S-PLUS*, Springer, New York (2000).

Jonathan Phillips obtained his BS in chemistry from Wheaton College (Illinois) and his MS in color science from RIT. He is currently pursuing his PhD in color science at RIT in the areas of high dynamic range imaging and material appearance. He has worked at Eastman Kodak Company since 1992, where he is a principal scientist in the Office of the Chief Technical Office. Phillips has worked as an environmental analytical chemist, a systems engineer on photographic film and magnetics, and a lead image scientist on commercialization of printers, displays, and cameras.



Peter Bajorski received his BS and MS in mathematics from the University of Wroclaw, Poland, and his PhD in mathematical statistics from the Technical University of Wroclaw, Poland. He has held positions at Cornell University, the University of British Columbia, Simon Fraser University, and the NY State Department of Transportation. Currently, he is associate professor in the Graduate Statistics Department at RIT. His research interests in imaging include statistical methods for remote sensing and color science. Other interests include multivariate statistical methods, regression techniques, and design of experiments. Dr. Bajorski is a member of ASA, SPIE, and IEEE.

Peter Burns is principal scientist with Carestream Health's Research and Innovation Labs. His work includes medical and dental image processing, and the development of imaging performance evaluation, analysis, and software tools. He also worked in this area at Eastman Kodak Company and Xerox, and has taught imaging courses at Kodak, SPIE, and IS&T Conferences, and at the Center for Imaging Science, RIT. He received his BS and ME in electrical engineering from Clarkson University and a PhD in imaging science from RIT.

Erin Fredericks received her BS in imaging and photographic technology in 2007 and MS in imaging science in 2009 from RIT. Her focus has been on preferred color reproduction. She is a color scientist at ITT Corporation.

Mitchell Rosen is a research professor in the Munsell Color Science Laboratory of the Center for Imaging Science at RIT and director of the Infinite Pixel Liberation Laboratory (iPixLab). He was color imaging editor of the *Journal of Imaging Science and Technology* from 2002 to 2007. His research interests include spectral and colorimetric color reproduction, computational optimization methods for color reproduction, color management, visual perception, and innovative imaging modalities.