



Document Color Usage Analysis

QualityLogic Research Note 032805

March 22, 2005

Table of Contents

1.	Overview	3
2.	Analysis Methodology.....	3
3.	Results	4
4.	Conclusions and Implications	5
5.	Appendix A.....	7
6.	Appendix B.....	9

1. Overview

QualityLogic develops a wide variety of real world test pages using popular software applications. These test pages are used for interoperability testing of drivers, printer system performance testing, and yield testing of printer systems. The distribution of colors used in these test pages has a significant impact on their effectiveness as test tools.

In order to better understand what real world color usage looks like, QualityLogic undertook a study to analyze 10,744 PDF files that had been randomly downloaded as part of the development of QualityLogic's PDF InteropAnalyzer product. These files were pulled from approximately 700 web sites focused on a wide variety of topics, yielding a representative sample of the "real world."

In order to ensure that these PDF files are actually documents users are likely to print, we imposed the following constraints on the files used in the analysis:

- Documents could not exceed 20 pages in length. We found that documents exceeding this length were typically catalogs or other references works which were very unlikely to be printed in their entirety by end users.
- Page size must be letter or A4. Documents with other sizes were typically content specifically designed for the web or other mediums.
- The documents must use at least one of the primary colors (CMY).

The result of these constraints netted 7,689 files for analysis.

2. Analysis Methodology

QualityLogic used the PA Coverage Expert to analyze the color coverage on each page. This tool uses GhostScript to generate a rasterized image of the page, then uses the raster image to calculate page coverage. We compared the reported page coverage analysis from this tool using two other methodologies, and we believe that the coverage reported by the PA tool is accurate enough to draw some general conclusions. The alternative methodologies evaluated include using the histogram feature of PhotoShop on color separations of the documents, as well as using a tool called APFill, which also used GhostScript as the basis for its analysis. See Appendix A for a report comparing all three tools.

The page coverage percentages for each PDF document were added to the PDF InteropAnalyzer database. The PDF InteropAnalyzer QueryBuilder utility was then used to extract what we think are useful facts from the data.

3. Results

The following is a summary of the results of this analysis.

- Average usage for the entire population of 7,689 files was as follows. As expected, there was a huge distribution of color usage on a document to document basis, with standard deviations between 3.4% and 5.2% for individual colors across the entire set of files.

	Cyan	Magenta	Yellow	Black	Sum of CMYK
Coverage	3.9%	4.3%	4.0%	4.5%	16.7%

- When looking at Toner usage across different PDF versions, it is clear that documents using more recent version of PDF appear to use more color. This could support the contention that the use of color in office documents is growing, although it may also be a function of more sophisticated early adopters using the latest versions of Acrobat.

PDF Version	# of Files	Cyan	Magenta	Yellow	Black	Sum of CMYK
1.2	2617	3.9%	4.2%	3.9%	4.4%	16.5%
1.3	1391	3.7%	4.1%	3.9%	4.4%	16.2%
1.4	1457	4.3%	4.8%	4.4%	4.7%	18.2%
1.5	191	4.7%	5.5%	4.8%	5.2%	20.3%

- There are significant differences in the amount of color used depending on the application that created the document. The following table sums the percentages of CMYK to provide an illustration of these differences:

Application Program	Sum of CMYK Coverage
FrameMaker	7.3%
Word	10.7%
Excel	14.5%
PageMaker	21.8%
CorelDraw	23.4%
QuarkXPress	25.2%
Illustrator	25.7%
InDesign	26.6%
PowerPoint	36.5%

- We also looked at page coverage as a function of document length.

Document Length	Sum of CMYK Coverage
1 page	18.4%
2 - 5 Pages	17.4%
6 - 20 pages	16.2%

- Attempts to look at image resolution as a function of page coverage did not result in any interesting conclusions. Appendix B contains the raw data used to form the summary tables above.

4. Conclusions and Implications

The information extracted from the PDF files provides QualityLogic with some useful metrics for the design of color test pages used in our interoperability and performance test tools.

Unexplored areas that might be of interest to our customers could include looking at color coverage in specific geographic markets or by specific industry segments. This can be done by collecting PDF files from specific regions or on web sites with specific topics. Another possible application of this data would be to mine the database for test pages with specific color coverage characteristics.

One conclusion that somewhat surprised us was the validation that the conventional five percent per color page coverage used for yield, performance

and reliability testing is a reasonable approximation of the real world as reflected by this analysis.

There remains the question of whether a set of files from the Internet is really representative of the actual printing patterns of end-users. While we have attempted to constrain the analysis to increase the probability that we are seeing pages likely to be printed, we need some independent comparisons before claiming that this analysis represents the real world of everyday printing.

5. Appendix A

The following tables compare the page coverage of five ISO yield test pages using three different methodologies. There are some surprisingly large differences in reported yield on a color-by-color basis between each of these methodologies, illustrating that attempting to calculate coverage without actually putting dots on paper is a somewhat inexact science. When looked at in the aggregate (sum of all four colors on a page), the variations are somewhat lessened. We believe the results from the PA tool are adequate to draw our general conclusions.

Page 1			
	PA Anal	APFill	PhotoShop
Cyan	0.21	0.1	0.21
Magenta	0.65	0.55	0.71
Yellow	1.25	1.15	1.35
Black	4.84	3.65	3.95
Sum	6.95	5.45	6.22

Page 2			
	PA Anal	APFill	PhotoShop
Cyan	3.45	3.57	3.78
Magenta	2.65	2.73	2.7
Yellow	2.65	2.73	2.81
Black	3.05	2.52	2.45
Sum	11.8	11.55	11.74

Page 3			
	PA Anal	APFill	PhotoShop
Cyan	2.79	2.28	3.6
Magenta	5.3	4.56	5.54
Yellow	5.38	4.56	5.58
Black	8.7	8.55	7.89
Sum	22.17	19.95	22.61

Page 4			
	PA Anal	APFill	PhotoShop
Cyan	2.41	2.24	2.68
Magenta	4.01	4.34	4.81
Yellow	3.97	3.78	4.21
Black	3.31	3.5	2.97
Sum	13.7	13.86	14.67

Page 5			
	PA Anal	APFill	PhotoShop
Cyan	10.88	12.32	10.78
Magenta	10.62	11.76	10.5
Yellow	9.63	10.64	10.79
Black	4.33	5.04	4.68
Sum	35.46	39.76	36.75

6. Appendix B

The following are the raw results extracted from the Query Analyzer when looking for page coverage trends.

7689 Records C:3.9 M:4.3 Y:4.0 B:4.5 Tot: 16.7

PDF Version

v1.2 2617 Records C:3.9 M:4.2 Y:3.9 B:4.4 Tot: 16.5

v1.3 3291 Records C:3.7 M:4.1 Y:3.9 B:4.4 Tot: 16.2

v1.4 1457 Records C:4.3 M:4.8 Y:4.4 B:4.7 Tot: 18.2

v1.5 191 Records C:4.7 M:5.5 Y:4.8 B:5.2 Tot: 20.3

Creators

CorelDraw	44 Records C:5.7 M:6.8 Y:5.5 B:5.4 Tot: 23.4
Excel	94 Records C:3.1 M:3.3 Y:3.3 B:4.8 Tot: 14.5
FrameMaker	66 Records C:1.4 M:1.7 Y:1.3 B:2.9 Tot: 7.3
Illustrator	158 Records C:7.0 M:6.2 Y:6.1 B:6.4 Tot: 25.7
InDesign	203 Records C:5.9 M:6.6 Y:7.1 B:7.0 Tot: 26.6
PageMaker	476 Records C:5.1 M:5.7 Y:5.6 B:5.4 Tot: 21.8
PowerPoint	51 Records C:11.5 M:11.4 Y:7.9 B:5.7 Tot: 36.5
QuarkXPress	1039 Records C:6.1 M:6.4 Y:6.8 B:5.9 Tot: 25.2
Word	2145 Records C:2.3 M:2.6 Y:2.2 B:3.6 Tot: 10.7

Number of pages

1 Page	1704 Records C:4.2 M:4.6 Y:4.5 B:5.2 Tot: 18.4
2 – 5 Pages	4174 Records C:4.0 M:4.4 Y:4.1 B:4.8 Tot: 17.4
6 – 20 Pages	1811 Records C:3.8 M:4.2 Y:3.8 B:4.3 Tot: 16.2

Images

Has Images	5436 Records C:4.6 M:5.1 Y:4.7 B:4.9 Tot: 19.3
1 -200 dpi	4753 Records C:4.8 M:5.3 Y:5.0 B:4.9 Tot: 20.0
201 – 400 dpi	1459 Records C:4.6 M:5.0 Y:4.6 B:5.1 Tot: 19.2
401 - 600 dpi	294 Records C:3.7 M:4.0 Y:4.3 B:5.3 Tot: 17.4
601 - 1000 dpi	179 Records C:4.0 M:4.3 Y:3.8 B:6.2 Tot: 18.4
1001 – 2000 dpi	115 Records C:4.7 M:5.4 Y:5.4 B:5.2 Tot: 20.7
2000 dpi plus	33 Records C:2.7 M:3.3 Y:4.0 B:5.8 Tot: 15.7