Legibility and Large-Scale Digitization

Jeremy York
November, 2008

Introduction

The large-scale digitization initiatives undertaken by Google, the Open Content Alliance (OCA), the Library of Congress, and others in recent years have raised significant questions for libraries and institutions seeking to provide better and faster access to their information resources. These range from debates over copyright and ownership, to considerations of the future of books and reading, to the development of standards and best practices for digitization that seek to balance competing desires for quality and quantity in digital output. This paper delves more deeply into one strain in these discussions that is of increasing importance as greater and greater amounts of content are delivered to users in, and from, digital formats: this is the presentation, and specifically the legibility of digital materials that are made available on computer screens, and on paper as print-on-demand products from digital files.

The issue of legibility has long been a concern in both arenas (on computer screens and in print), but has reemerged with the large-scale digitization of print collections because of an increased interest in the capture of grey scale and color page images. Conventional practice has been to use 600 dpi TIFF “Group 4” bitonal imaging with lossless compression for printed text when it is possible (when scans meet requirements for image performance (OCR) and quality) and some form of lossy compression such as JPEG when it is not (when illustrations or handwriting appear in the content to be scanned, for example).[1] Although JPEG compression is accompanied by losses in visual quality, the size of uncompressed TIFF masters for grey scale and color images can be prohibitively large.

In the year 2000, a new standard for the capture of master digital files was introduced that is gaining traction as an alternative to these methods. The standard, known as JPEG 2000, is able to capture lossless and visually lossless compressed digital image files that are significantly smaller than traditional TIFF files for grey scale and color images. Bitonal JPEG 2000 images do not stack up as well against their bitonal TIFF counterparts as far as file size is concerned,

[1] Chapman, Stephen, Duplouy, Laurent, Kunze, John, Blair, Stuart, Abrams, Stephen, Lupovici, Catherine, Jensen, Ann, Johnston, Dan (2006). “Page Image Compression for Mass Digitization”; *IS&T Archiving 2007 Conference Proceedings*, p.2. This study will be discussed further below.

(they are 30% larger on average than 600dpi TIFF G4 files), but this difference is judged increasingly to be within reason.

A study done by Google and OCA partners (Harvard and the University of California Berkeley) and the Bibliotheque Nationale de France in 2006 found that visually lossless JPEG 2000 images performed optimally in meeting the principle requirements of mass digitization mandated by today's technological and economic constraints. These requirements, as stated in the study, are to:

- enable very fast scanning of bound volumes (to reduce costs associated with human handling time);
- yield page image masters adequate for OCR and production of images with readable (legible) content when rendered as soft- and hard-copy outputs; and
- result in small file sizes, with the two-fold benefit of speeding up online transfer (ingest and access) and minimizing per unit (page/volume) storage costs.[2]

Partners in the study showed that JPEG 2000 scans of "'marginal' or better quality[3] could be produced with file sizes averaging between 181-225KB for text pages and 268-372KB for non-text pages. These represent significant savings over multi-megabyte grayscale and color master files that TIFF produces, with additional savings in processing overhead for JPEG 2000 images when compared with conventional practices for generating access copies of digital files from preservation masters.

Findings such as these illustrate why JPEG 2000 is emerging as the best option for capturing and displaying grayscale and color images in large-scale digitization initiatives. Concerns about legibility have arisen, however, as more institutions engaging in these projects are selecting JPEG 2000 as the default format for all page image scans, and not only those where unacceptable degradation of the image with 600 dpi TIFF G4 is observed. The result, besides increasing file sizes for master copies of individual volumes (average file sizes for bitonal TIFF images are between 105 and 120KB per page, compared with 181-225KB for JPEG 2000[4]), is that in some cases the contrast between the foreground text and background of these color and grayscale images is low, making the text difficult to read.

---

[2] Chapman, Duplouy, Kunze, Blair, Abrams, Lupovici, Jensen, Johnston, (2006), p.1.

[3] Judgments of quality were based on users' ratings of file samples on a scale of Perfect, Acceptable, Marginal, and Unacceptable when compared with uncompressed TIFF files.

[4] Chapman, Duplouy, Kunze, Blair, Abrams, Lupovici, Jensen, Johnston, (2006), pp.2,4.

The W3C consortium, in its new Web Content Accessibility Guidelines 2.0 candidate recommendations, includes a guideline to "Make it easier for users to see and hear content including separating foreground from background."[5] This guideline includes recommended contrast ratios[6] for text and images of text against their backgrounds at minimum and enhanced levels. The minimum level calls for a ratio of at least 5:1[7] between background and text, and the enhanced level for 7:1. These levels are meant to ensure accessibility to users with moderately low vision (approximately 20/40 vision) and low vision (20/80), respectively.[8] The minimum recommended contrast ratio for users with normal vision is 3:1.[9]

A sample of three of the most downloaded PDF books from the Open Content Alliance[10] on October 27, 2008, showed that while the greatest contrast on randomly selected pages was in compliance with both the minimum and enhanced levels of the W3C candidate recommendation, average and lighter contrasts were on the border line of compliant, or not compliant. The results of the samples are shown in the figures and tables below. The AA level corresponds to a 5:1 ratio and AAA to a 7:1 ratio. For comparison purposes, the contrast ratio for perfectly black text (hexadecimal value #000000) on a white background (#ffffff) is 21:1.

---

[5] Guideline 1.4. Web Content Accessibility Guidelines 2.0, W3C Candidate Recommendation 30 April 2008, http://www.w3.org/TR/WCAG20/#visual-audio-contrast-contrast (accessed October 27, 2008).

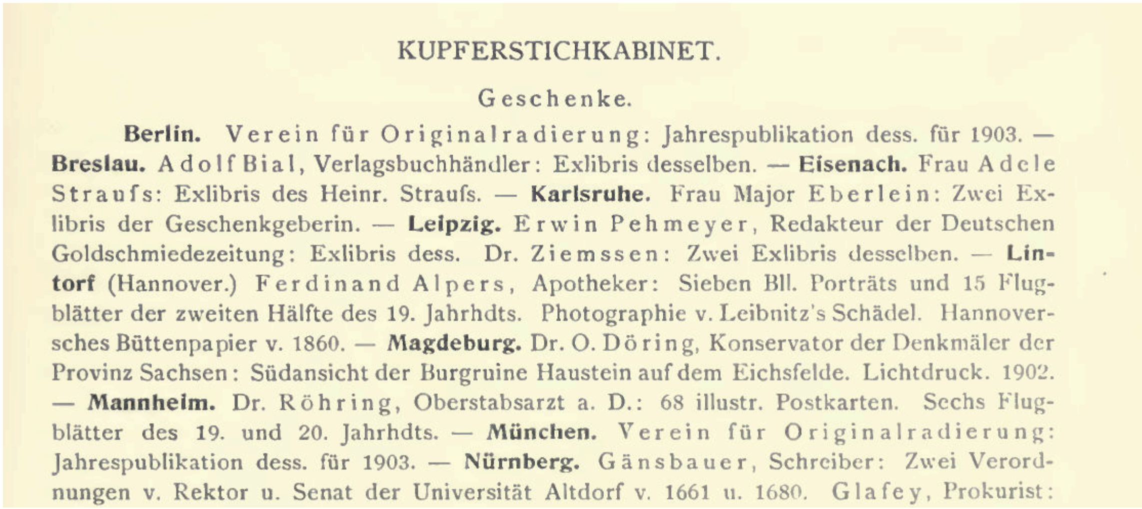
[6] The differences in contrast are differences in relative luminance. See the procedure for calculating relative luminance at http://www.w3.org/TR/WCAG20-GENERAL/G17.html.

[7] This item is listed as "at risk" with the possibility of changing the recommendation to 4.5:1 or 4:1 if it proves too restricted.

[8] W3C Consortium. (2008) "Contrast (Minimum)", *Understanding WCAG 2.0: A Guide to Understanding and Implementing WCAG 2.0, W3C Working Draft*, http://www.w3.org/TR/2008/WD-UNDERSTANDING-WCAG20-20080430/visual-audio-contrast-contrast.html (accessed October 27, 2008).

[9] See ISO-9241-3 and ANSI-HFES.

[10] http://www.archive.org/details/opencontentalliance (accessed November 10, 2008).



KUPFERSTICHKABINET.

Geschenke.

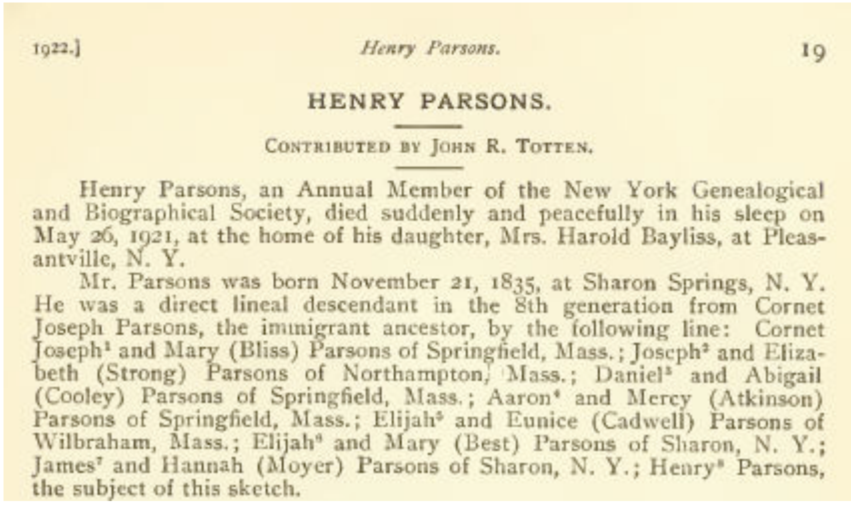
Berlin. Verein für Originalradierung: Jahrespublikation dess. für 1903. — **Breslau.** Adolf Bial, Verlagsbuchhändler: Exlibris desselben. — **Eisenach.** Frau Adele Straufs: Exlibris des Heinr. Straufs. — **Karlsruhe.** Frau Major Eberlein: Zwei Exlibris der Geschenkgeberin. — **Leipzig.** Erwin Pehmeyer, Redakteur der Deutschen Goldschmiedezeitung: Exlibris dess. Dr. Ziemssen: Zwei Exlibris desselben. — **Lintorf** (Hannover.) Ferdinand Alpers, Apotheker: Sieben Bll. Porträts und 15 Flugblätter der zweiten Hälfte des 19. Jahrhdts. Photographie v. Leibnitz's Schädel. Hannoversches Büttenpapier v. 1860. — **Magdeburg.** Dr. O. Döring, Konservator der Denkmäler der Provinz Sachsen: Südansicht der Burgruine Haustein auf dem Eichsfelde. Lichtdruck. 1902. — **Mannheim.** Dr. Röhring, Oberstabsarzt a. D.: 68 illustr. Postkarten. Sechs Flugblätter des 19. und 20. Jahrhdts. — **München.** Verein für Originalradierung: Jahrespublikation dess. für 1903. — **Nürnberg.** Gänsbauer, Schreiber: Zwei Verordnungen v. Rektor u. Senat der Universität Altdorf v. 1661 u. 1680. Glafey, Prokurist:

Figure I. Germanisches Nationalmuseum Nürnberg (1904-06). *Anzeiger des Germanischen National-Museums (1886)*; Verlagseigentum des germanischen Museums, Nürnberg, p.VII.[11]

	Background	Foreground	Ratio	
Greatest contrast (Bold text)	#f5f8cb	#46404c	9.1:1	Pass at AAA
Average contrast	#fbf9d3	#534f43	7.6:1	Pass at AAA
Lightest contrast	#fffcd8	#827E7D	3.9:1	Fail at AA and AAA

Table I. Ratios of luminance contrast for text in different areas of page VII – *Anzeiger des Germanischen National-Museums (1886).*

[11] http://www.archive.org/details/anzeigerdesgerma1904to06germiala (accessed October 27, 2008).



1922.] *Henry Parsons.* 19

HENRY PARSONS.

CONTRIBUTED BY JOHN R. TOTTEN.

Henry Parsons, an Annual Member of the New York Genealogical and Biographical Society, died suddenly and peacefully in his sleep on May 26, 1921, at the home of his daughter, Mrs. Harold Bayliss, at Pleasantville, N. Y.

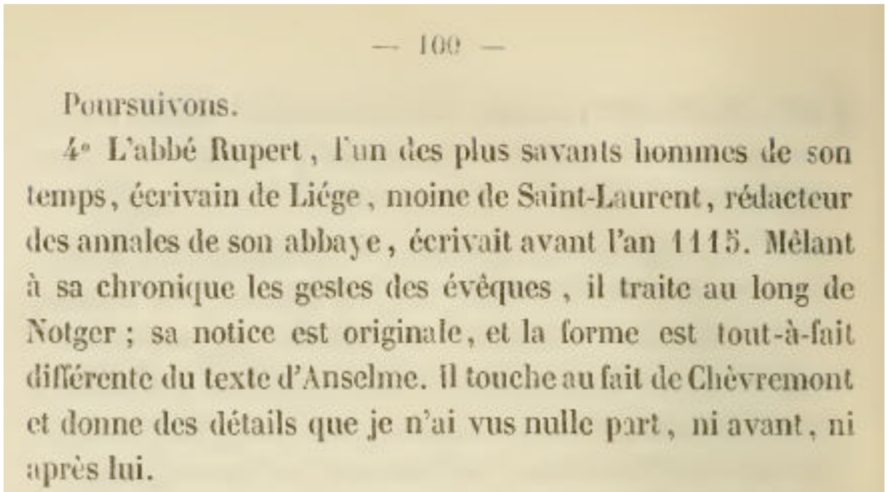
Mr. Parsons was born November 21, 1835, at Sharon Springs, N. Y. He was a direct lineal descendant in the 8th generation from Cornet Joseph Parsons, the immigrant ancestor, by the following line: Cornet Joseph[1] and Mary (Bliss) Parsons of Springfield, Mass.; Joseph[2] and Elizabeth (Strong) Parsons of Northampton, Mass.; Daniel[3] and Abigail (Cooley) Parsons of Springfield, Mass.; Aaron[4] and Mercy (Atkinson) Parsons of Springfield, Mass.; Elijah[5] and Eunice (Cadwell) Parsons of Wilbraham, Mass.; Elijah[6] and Mary (Best) Parsons of Sharon, N. Y.; James[7] and Hannah (Moyer) Parsons of Sharon, N. Y.; Henry[8] Parsons, the subject of this sketch.

Figure 2. New York Genealogical and Biographical Society (1977). *The New York genealogical and biographical record (1870) Index;* New York Genealogical and Biographical Society, New York, p.19.[12]

	Background	Foreground	Ratio	
Greatest contrast (excluding header)	#ffffd3	#594f34	7.9:1	Pass at AAA
Average contrast:	#fbf3ce	#61583b	6.3:1	Pass at AA
Lightest contrast	#ffffdc	#8D8561	3.6:1	Fail at AA and AAA

Table II. Ratios of luminance contrast for text in different areas of page 19 – *The New York genealogical and biographical record (1870) Index*.

[12] http://www.archive.org/details/newyorkgenealogi108newy (accessed October 27, 2008).


— 100 —

Poursuivons.

4° L'abbé Rupert, l'un des plus savants hommes de son temps, écrivain de Liége, moine de Saint-Laurent, rédacteur des annales de son abbaye, écrivait avant l'an 1115. Mêlant à sa chronique les gestes des évêques, il traite au long de Notger ; sa notice est originale, et la forme est tout-à-fait différente du texte d'Anselme. Il touche au fait de Chèvremont et donne des détails que je n'ai vus nulle part, ni avant, ni après lui.

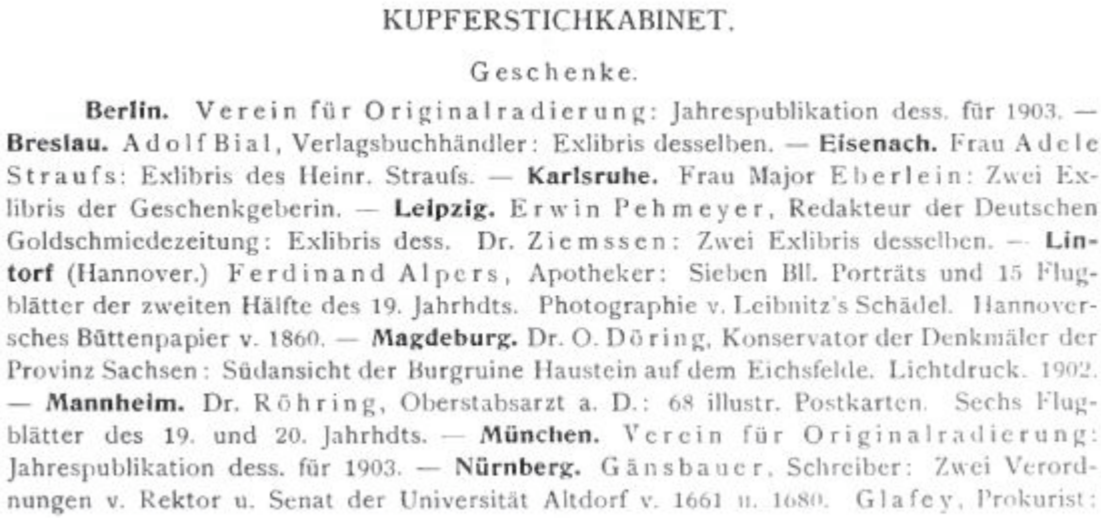
Figure 3. Institut archéologique liégeoise (1852-1901). *Bulletin de l'Institut archéologique liégeoise (1852)*; Maison Curtius, Liège, p.100.[13]

	Background	Foreground	Ratio	
Greatest contrast	#dbd1ae	#504529	6.2:1	Pass at AA
Average contrast	#e5dcbd	#564d2e	6.1:1	Pass at AA
Lightest contrast	#d1c9a5	#473d24	4.0:1	Fail at AA and AAA

Table III. Ratios of luminance contrast for text in different areas of page 100 – *Bulletin de l'Institut archéologique liégeoise (1852)*.

OCA does make black and white versions of their books available in addition to the color scans, but judging by their appearance and quality these versions are created as derivatives of the color page image scans, and not scanned separately. The results of the same tests performed on the black and white volumes are given below.

[13] http://www.archive.org/details/bulletindelinsti01inst (accessed October 27, 2008).



KUPFERSTICHKABINET.

Geschenke.

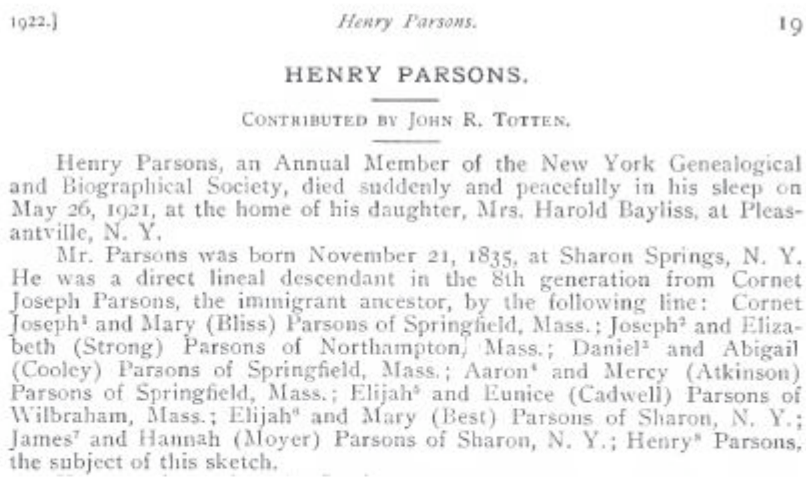
Berlin. Verein für Originalradierung: Jahrespublikation dess. für 1903. — **Breslau.** Adolf Bial, Verlagsbuchhändler: Exlibris desselben. — **Eisenach.** Frau Adele Straufs: Exlibris des Heinr. Straufs. — **Karlsruhe.** Frau Major Eberlein: Zwei Exlibris der Geschenkgeberin. — **Leipzig.** Erwin Pehmeyer, Redakteur der Deutschen Goldschmiedezeitung: Exlibris dess. Dr. Ziemssen: Zwei Exlibris desselben. — **Lintorf** (Hannover.) Ferdinand Alpers, Apotheker: Sieben Bll. Porträts und 15 Flugblätter der zweiten Hälfte des 19. Jahrhdts. Photographie v. Leibnitz's Schädel. Hannoversches Büttenpapier v. 1860. — **Magdeburg.** Dr. O. Döring, Konservator der Denkmäler der Provinz Sachsen: Südansicht der Burgruine Haustein auf dem Eichsfelde. Lichtdruck. 1902. — **Mannheim.** Dr. Röhring, Oberstabsarzt a. D.: 68 illustr. Postkarten. Sechs Flugblätter des 19. und 20. Jahrhdts. — **München.** Verein für Originalradierung: Jahrespublikation dess. für 1903. — **Nürnberg.** Gänsbauer, Schreiber: Zwei Verordnungen v. Rektor u. Senat der Universität Altdorf v. 1661 u. 1680. Glafey, Prokurist:

Figure 4. Germanisches Nationalmuseum Nürnberg (1904-06). *Anzeiger des Germanischen National-Museums (1886)*; Verlagseigentum des germanischen Museums, Nürnberg, p.VII [Black and White].[14]

	Background	Foreground	Ratio	
Greatest contrast (Bold text)	#ffffff	#1d1d1d	16.9:1	Pass at AAA
Average contrast:	#ffffff	#5b5b5b	6.8:1	Pass at AA
Lightest contrast	#ffffff	#aaaaaa	2.3:1	Fail at AA and AAA

Table IV. Ratios of luminance contrast for text in different areas of page VII – *Anzeiger des Germanischen National-Museums (1886)* [Black and White].

[14] http://www.archive.org/details/anzeigerdesgerma1904to06germiala (accessed October 27, 2008).


1922.] Henry Parsons. 19

HENRY PARSONS.

CONTRIBUTED BY JOHN R. TOTTEN.

Henry Parsons, an Annual Member of the New York Genealogical and Biographical Society, died suddenly and peacefully in his sleep on May 26, 1921, at the home of his daughter, Mrs. Harold Bayliss, at Pleasantville, N. Y.

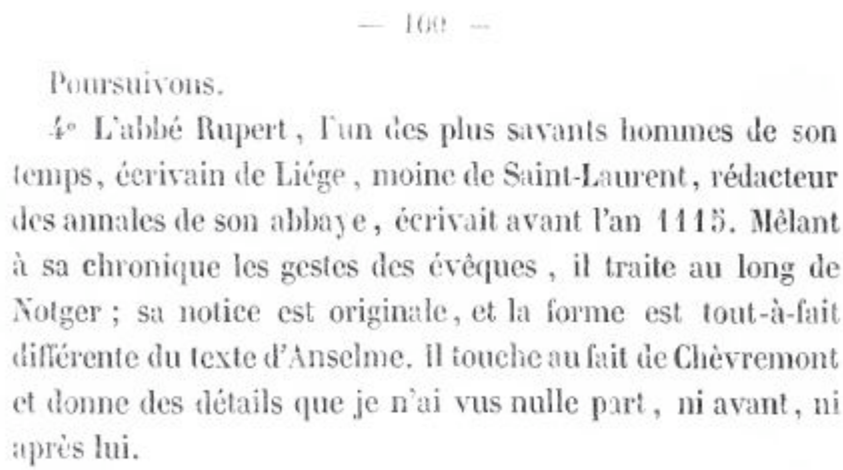
Mr. Parsons was born November 21, 1835, at Sharon Springs, N. Y. He was a direct lineal descendant in the 8th generation from Cornet Joseph Parsons, the immigrant ancestor, by the following line: Cornet Joseph[1] and Mary (Bliss) Parsons of Springfield, Mass.; Joseph[2] and Elizabeth (Strong) Parsons of Northampton, Mass.; Daniel[3] and Abigail (Cooley) Parsons of Springfield, Mass.; Aaron[4] and Mercy (Atkinson) Parsons of Springfield, Mass.; Elijah[5] and Eunice (Cadwell) Parsons of Wilbraham, Mass.; Elijah[6] and Mary (Best) Parsons of Sharon, N. Y.; James[7] and Hannah (Moyer) Parsons of Sharon, N. Y.; Henry[8] Parsons, the subject of this sketch.

Figure 5. New York Genealogical and Biographical Society (1977). *The New York genealogical and biographical record (1870) Index;* New York Genealogical and Biographical Society, New York, p.19 [Black and White].[15]

	Background	Foreground	Ratio	
Greatest contrast (excluding header)	#ffffff	#3f3f3f	8.1:1	Pass at AAA
Average contrast:	#ffffff	#7b7b7b	4.2:1	Fail at AA and AAA
Lightest contrast	#ffffff	#a1a1a1	2.6:1	Fail at AA and AAA

Table V. Ratios of luminance contrast for text in different areas of page 19 – *The New York genealogical and biographical record (1870) Index* [Black and White].

15 http://www.archive.org/details/newyorkgenealogi108newy (accessed October 27, 2008).


— 100 —

Poursuivons.

4° L'abbé Rupert, l'un des plus savants hommes de son temps, écrivain de Liége, moine de Saint-Laurent, rédacteur des annales de son abbaye, écrivait avant l'an 1115. Mêlant à sa chronique les gestes des évêques, il traite au long de Notger ; sa notice est originale, et la forme est tout-à-fait différente du texte d'Anselme. Il touche au fait de Chèvremont et donne des détails que je n'ai vus nulle part, ni avant, ni après lui.

Figure 7. Institut archéologique liégeoise (1852-1901). *Bulletin de l'Institut archéologique liégeoise (1852)*; Maison Curtius, Liège, p.100 [Black and White].[16]

	Background	Foreground	Ratio	
Greatest contrast	#ffffff	#37363b	12:1	Pass at AAA
Average contrast	#ffffff	#5c5c5c	6.7:1	Pass at AA
Lightest contrast	#ffffff	#949494	3.0:1	Fail at AA and AAA

Table VII. Ratios of luminance contrast for text in different areas of page 100 – *Bulletin de l'Institut archéologique liégeoise (1852)* [Black and White].

In all of these samples, PDF files were downloaded from the Internet and screen captures at 100% size where opened in Photoshop. The color-picker feature was used to find the hexadecimal text and background colors for areas that appeared to have high, "average", and low contrast. The darkest possible text color was taken (color varied significantly within individual letters), and the background color immediately surrounding it.

These methods are somewhat suspect as a means of making determinations about the overall legibility of the texts. However, the point to be made follows not from the exact contrast ratios that were recorded, but their general range. The large-scale digitization projects that are currently underway will determine to a large degree how texts are accessed on computer screens, in print (as print-on-demand hard copies), and through next generation media in

[16] http://www.archive.org/details/bulletindelinsti01inst (accessed October 27, 2008).

coming years. With so much at stake, it is essential to ensure that technological efficiency and convenience are not achieved at the expense of end-user experience and accessibility.

This paper does not attempt to advocate one file format over another or one method of processing over another. These decisions are best left to institutions, which have the best knowledge of their own materials and technology. The purpose is to stress the importance of taking decisions about digital capture and delivery that, while satisfying technological and economic demands, 1) prioritize accessibility of content to users with a wide variety of needs, in a variety of formats (including those not yet instantiated) and 2) are based on standards and findings from legibility research.

Standards for screen legibility have been introduced above and are fairly well-known and documented on the Web. A number of websites and applications exist that test color combinations to determine the relative luminance contrast between textual and background colors,[17] and guidelines for color usage and display are available from entities such as NASA's Color Usage Research Lab[18] and, of course, the W3C consortium. The W3C consortium cites the research of Kenneth Knoblauch and Aries Arditi (1991, 1994, 1996 and 2004) in making its candidate recommendation on color contrast.[19] Significant work in this area has also been done more recently by Gordon Legge and others.[20] These studies have demonstrated the importance of luminance contrast (and relative unimportance, on the other hand, of color contrast) for legibility on computer screens, showing significant losses in legibility below contrast ratios of 3:1 for readers with normal vision, and 5:1 and 7:1 for readers with progressively lower levels of vision.

---

[17] Some of these are listed at http://www.456bereastreet.com/archive/200709/10_colour_contrast_checking_tools_to_improve_the_accessibility_of_your_design/ and http://www.w3.org/TR/2008/WD-UNDERSTANDING-WCAG20-20080430/visual-audio-contrast7.html (all accessed October 27, 2008).

[18] http://colorusage.arc.nasa.gov/index.php (accessed October 27, 2008).

[19] Knoblauch, K., Arditi, A., & Szlyk, J. (1991) Effects of chromatic and luminance contrast on reading. *Journal of the Optical Society of America*, 8, 428-439; Arditi, A. and Knoblauch, K. (1994). Choosing effective display colors for the partially-sighted. Society for Information Display International Symposium Digest of Technical Papers, 25, 32-35; Arditi, A. and Knoblauch, K. (1996). Effective color contrast and low vision. In B. Rosenthal and R. Cole (Eds.) Functional Assessment of Low Vision. St. Louis, Mosby, 129-135; Arditi, A. and Faye, E. (2004). Monocular and binocular letter contrast sensitivity and letter acuity in a diverse ophthalmologic practice. Supplement to Optometry and Vision Science, 81 (12S), 287.

[20] Legge, Gordon E. (2006). Psychophysics of Reading in Normal and Low Vision, Routeledge. See also Zuffi, Silvia, Brambilla, Carla, Beretta, Giordano, Scala, Paolo (2007) "Human Computer Interaction: Legibility and Contrast"; *Proceedings of the 14th International Conference on Image Analysis and Processing*, IEEE Computer Society, Washington, D.C.; Gradisar, M., Iztok. H., Turk, T. (2006). "Factors Affecting the Readability of Colored Text on Computer Displays"; *28th International Conference on Information Technology Interfaces*, 2006.

Standards for print legibility are also available on the Web, through organizations such as the Canadian National Institute for the Blind,[21] the Canadian Public Health Institute,[22] and Lighthouse International,[23] but supporting research is more difficult to find. This is because many of the studies on print legibility were done before 1970 (particularly between the 1920s and 1960s) and are available only through restricted journal access on the Internet or in print (as opposed to more broad availability through general web search). Furthermore, the problems of print legibility were largely solved by the time computers were introduced and screen legibility became an issue. By this time, principles and practices of publishing were very familiar to traditional advertisers, designers, and printers and interest in legibility shifted to the functional efficiency and, particularly after the information of the World Wide Web, new information distribution capabilities of reading and working in the new medium.

Opinions and research about legibility on computers and the Web are ongoing, but rather than concentrate on present trends to discuss legibility in large-scale digitization projects, this paper looks to the past to inform present and future decisions about how we preserve and present our accumulated knowledge. There are several reasons for this approach. The first is that space does not allow for an in-depth treatment of both screen and print usability. The second is that the "answer" is basically the same in both cases – computers emit light and printed materials reflect light, but legibility in both cases is determined by the contrast (in relative luminance for computer screens and reflectance or brightness contrast for print) between text and background. The third, and possibly most important reason, is that paper is and will likely remain a preferred reading format for readers of all ages and needs in all parts of the world. Investigating the processes and research that have formed the printed world as we know it can inform our present and future decisions about the capture and delivery of digital content, and ensure that it is accessible to the widest possible audience.

---

[21] http://www.cnib.ca/en/services/accessibility/text/clearprint/CNIB%20Clear%20Print%20Guide.pdf (accessed October 27, 2008). See also the CNIB website at http://www.cnib.ca/, with resources at http://www.cnib.ca/en/services/accessibility/text/clearprint/Default.aspx.

[22] http://www.nlhp.cpha.ca/. See http://www.nlhp.cpha.ca/Labels/seniors/english/GoodMed-E.pdf (accessed October 27, 2008).

[23] See guidelines by Aries Arditi at http://www.lighthouse.org/accessibility/legible/ (accessed October 27, 2008).

Print Legibility

The question of which textual and background colors make the best combination for reading printed materials has been studied extensively over the last two centuries. While some of these studies have been poorly documented and therefore difficult to repeat or verify, the collective body of research in this area points to a single overarching conclusion: the greatest legibility of printed materials is achieved through the greatest difference in "brightness contrast" between printed text and its background.

Brightness contrast, also known as reflectance, is the ratio of the amount of light reflected by a surface to the amount of light striking the surface. A perfectly reflecting surface would have a reflectance of 1 or 100% and a perfectly non-reflecting surface would have a reflectance of 0 or 0%.[24] Evidence from the studies summarized below suggests that contrasts between backgrounds with a high reflectance such as white, which M. A. Tinker[25] calculates at over 70%, and text with a low reflectance such as black, at 3 to 4%, create the best environment for reading printed materials.[26] It has been found, moreover, that the difference in reflectance is the determinant of optimal legibility regardless of the colors that are used. For instance, if a shade of green is printed on shade of white, as long as the reflectance difference between the two is above 65% (according to Tinker), there will not be a significant difference in legibility between this and other combinations with similarly high differences in reflectance.[27]

Background

Concerns about aspects of printing such as the typeface, layout, and kind and quality of paper are as old as printing itself, but attention to the legibility of text in particular began in the late eighteenth and early nineteenth centuries.[28] Before this time, as one early twentieth-century investigator, R. L. Pyke, states, "it was the aesthetic aspect with which printers were most deeply concerned." Other factors also came into play in decisions about printing, such as the costs of

---

[24] Sanders. Mark S. and McCormick, Ernest J. (1993). *Human Factors In Engineering and Design (7th ed.),* McGraw-Hill, Inc., New York. Pages 516-519 provide background on how measurements of light are calculated, including reflectance.

[25] Miles Tinker was a leading researcher in legibility studies from the mid-1920s to the mid-1970s. The work of Tinker, and his colleague Donald Paterson, was a driving force behind the standardization of the print industry in the United States. See (Stone, Deborah (1997). *The Legibility of Text on Paper and Laptop Computer: A Multivariable Approach*. Dissertation, p. 38,139).

[26] Tinker, M. A. (1963). *Legibility of Print*; Iowa State University Press, Ames, Iowa, p.147.

[27] Tinker (1963), p.150.

[28] Pyke, R. L. (1926). "Report on the Legibility of Print"; *Medical Research Council*, Special Report Series, p.6.

printing certain sizes of font and the capabilities of existing printing technologies. However, although the combined result of their efforts often "achieved legibility" the primary concern of printers during this time was aesthetic appearance.[29]

This began to change in the nineteenth century as interest grew in psychology, physiology, education, and advertising, but it was a slow transition to more rigorous investigations of the factors influencing the legibility of print. In 1926, R. L. Pyke prepared a report for the Committee Upon The Legibility of Type in Great Britain in which he was highly critical of the poor methodology and lack of consistent criteria that was used in legibility studies to that time -- studies that were often engineered to support researchers' opinions or based on observations in the absence of consistent controls. In an attempt to consolidate and categorize past legibility research and provide a foundation for future experimentation, Pyke identified eighteen sub-topics, or categories of study that made up the field of legibility, and gave a summary of the work done in each of them. They were:

1. Contrast of thickness and thinness [of letters]
2. Criterion of legibility
3. Definition of legibility
4. Faces of type
5. Illumination
6. Indentation
7. Leading
8. Legibility of letters
9. Length of line
10. Margin
11. Paper and ink
12. Projectors
13. Punctuation
14. Serifs
15. Size of type
16. Spacing
17. 'The Ideal Type' [for typeface]
18. Thickness of limbs

These categories[30] are still relevant today and, for the purposes of this paper, help to locate research that has been done involving combinations of textual and background colors in the context of legibility studies more generally (according to Pyke's schema, it would fall under studies of Paper and Ink). A large number of factors come together to influence the overall legibility of text, and the current paper is a review of work done in only one of these.

---

[29] Pyke, p.6-7.
[30] Pyke, p.9.

Legibility v. Readability

Before continuing further, a note should be made about the third sub-topic in Pyke's list, the "Definition of legibility", and the relationship of legibility to readability, another term commonly used to describe printed materials today.

It has been a significant challenge in legibility studies, and one that is still a source of confusion today, to derive a coherent and unified definition of legibility. In general, it has always been agreed that legibility refers to the physical characteristics of text and figures presented on a page. Disagreement has occurred, however, over whether legibility refers more specifically to the ability to distinguish characters from one another, the ability to perceive characters, to easily read them, or to understand the meaning they are trying to convey.

In *Using Type,* published in 1996, Aernout de Beaufort Wijnholds defines legibility as "the attribute of alphanumeric characters that makes it possible for each one to be identifiable from others."[31] When applied to a body of text, he says, legibility refers to how easily individual characters can be grouped into words that are perceived to form a meaningful sentence.[32] Other researchers, however, such as Zachrisson[33] and Tinker[34] take a broader view of legibility that includes the ease with which a text is read and reading comprehension. Tinker, who did research on legibility from the 1920s through the 1970s (see note 24 above), states:

> Optimal legibility of print…is achieved by a typographical arrangement in which shape of letters and other symbols, characteristic word forms and all other typographical factors such as type size, line width, leading, etc., are coordinated to produce comfortable vision and easy and rapid reading with comprehension.[35]

There are two main sources of confusion between the terms legibility, whether narrowly or broadly defined, and readability. The first has to do with the fact that in 1940, the term readability began to be used by some writers nearly

---

[31] Sanders, M.S. and McCormick, E.J. (1993): *Human Factors in Engineering and Design*; McGraw-Hill, New York. Cited in de Beaufort Wijnholds, Aernout, Harm J. Zwaga, supervisor: *Using Type: The Typographer's Craftsmanship and the Ergonomist's Research;* Utrecht University, Netherlands. <http://www.plainlanguagenetwork.org/type/typelomx.htm>
[32] Sanders and McCormick.
[33] Zachrisson B. (1965): Studies in Legibility of Printed Text; Almqvist & Wiksell, Stockholm, p. 95.
[34] Tinker (1963) p.7-8.
[35] Tinker (1963), p.8.

synonymously with legibility.[36] Soon afterward, new studies in “readability” began that were concerned not so much with the perception of characters or groups of characters themselves, but the ease with which the informational content conveyed by those characters could be understood. These studies, utilizing “readability formulas” and “readability surveys”, were designed to measure the relative difficulty of the vocabulary, sentence structures, and abstract ideas that were used in a text, as well as tables, footnotes, and formatting that were used.[37] Within a few years, the word “readability” referred to two different, though related, areas of reading research.

The second source of confusion is that although legibility and readability refer to different regions of the reading spectrum, some of the same criteria, such as speed of reading and reader fatigue, are used in measures of each. Because the topic of this paper relates more closely to legibility than to readability (as readability is currently understood), the remainder of the discussion will focus on legibility, using the broader definition given by Tinker and Zachrission in particular. This definition and the criteria it is based upon (discussed below) provide a framework that is most inclusive of, and informative about, research relating to the study of background and textual colors.

Research

Criteria: How Legibility of Print is Measured

Succinct definitions of legibility (such as those of de Beaufort Wijnholds and Tinker above) are frequently cited, but legibility is often defined in practice by the criteria and methodologies that are used to investigate it. In his review of legibility research from 1825 to 1926, Pyke surveyed over one hundred studies and discovered fifteen different methods employed by researchers for measuring legibility. The methods he described were:

> …measurement by speed of reading (by the time threshold and amount read), the distance threshold (direct and peripheral), ‘eye-span’, ‘illumination threshold’, focus threshold, fatigue, number of eye-pauses, number of eye-refixations, regularity of eye-movements, reading rhythm, ‘legibility coefficient’, ‘specific legibility’, size of letters, by ‘judgment of the trained human eye’, and by aesthetic merits [as judged by subjects of the study].[38]

[36] Tinker (1963), p.8.
[37] de Beaufort Wijnholds, Aernout.
[38] Pyke, p.11.

In *Legibility of Print,* published nearly forty years later in 1963, Tinker presented a more condensed list of investigative criteria, representative of those most commonly employed. They included:

1. *Speed of perception*
   The speed and accuracy with which characters can be perceived in a short period of exposure.
2. *Perceptibility at a distance*
   The distance from the eyes (sometimes using an apparatus) at which characters can be accurately perceived.
3. *Perceptibility in Peripheral Vision*
   The distance from a given "fixation point" at which a character can be accurately perceived.
4. *Visibility*
   A measure of the point at which characters can be perceived when viewed through a visual apparatus that uses rotating filters to obscure and clarify those characters.
5. *The Reflex Blink Technique*
   Frequency of blinking when reading text with different typographical characteristics.
6. *Rate of work* [includes such measures as "speed of reading, amount of reading completed in a set time limit, time taken to find a telephone number, time taken to look up a power or root in mathematical tables, and work output in a variety of situations which involve visual discrimination."]
   A measure of the speed of reading, controlling for comprehension.
7. *Eye Movements*
   Measure of the movements of the eyes when reading, using methods such as corneal reflection and electrical signals.
8. *Fatigue in Reading*
   Has not been demonstrated to be a valid method for measuring legibility (see below).[39]

Pyke noted in his report the haphazard and inconsistent way these methodologies had been applied in the experiments he reviewed (some favoring the distance at which characters could be perceived as the best measure of

[39] Tinker (1963), pp.5-7.

legibility, for example, and others the speed at which they could read).[40] By the time *Legibility of Print* was published, however, it was understood that no single one of these methods (or criteria, depending on how they are described) was adequate for measuring legibility in all of its aspects. Each had to be understood and considered on its own merits as contributing to a broader notion of legibility. As Tinker says, "Some techniques supplement others to give a more complete picture of the legibility, while other techniques are limited to specific situations such as legibility of isolated characters."[41]

While this may seem to add more confusion the question of how legibility can be fully defined and measured, it helps to focus our discussion of research relating to combinations of textual and background colors. Aside from early investigations that were largely based on "casual observation",[42] research in this area was performed primarily on the basis of three criteria: 1) Speed of Perception; 2) Perceptibility at a Distance, and 3) Speed of Reading (under "Rate of work" in Tinker's list above).[43] These criteria are also those evaluated by Tinker to be most useful in measuring the effects of brightness contrast on legibility.

The remaining four criteria, and a fifth mentioned by Pyke, aesthetic appeal, have to a lesser degree been employed, and will be mentioned in conjunction with experiments below as they occur. To provide some initial context, however: Tinker found measurements of Visibility to be related to those of Speed of Perception and Perceptibility at a Distance, and measures of Eye Movement to be a valuable supplement for evaluating reading performance (Rate of work). The Reflex Blink Technique, most notably employed by Luckiesh, has been found by Tinker and others to be a largely unreliable and invalid method of investigation.[44] As regards Fatigue in Reading, although much research has been devoted to this area, sufficient methods for measuring its relation to legibility have not been found.[45]

Note: The experimental examples given below are not meant to be comprehensive, but to represent significant work that has been done (to varying degrees of scientific rigor) and approaches that have been taken in investigating

---

[40] Pyke, p.11.
[41] Tinker (1963), p.29.
[42] Tinker (1963) p.128.
[43] Tinker (1963). In Chapter 2, pp. 9-31, Tinker gives a description and evaluation of each of the eight criteria listed above. Of these,
[44] Tinker (1963), p. 17-19.
[45] Tinker (1963), p.20. This is due to the large number of factors involved in determining fatigue. See also Carmichael, L., and Dearborn, W. F. (1947). *Reading and Visual Fatigue*; Houghton Mifflin Co., Boston, pp. 206-451. Cited in Tinker, 1963.

legibility in general, and the question of contrast between text and background in particular. The works of Pyke, Tinker (1963), and Zachrisson (1965) taken together, list most of the experiments in this area dating from 1827.

Speed of Perception

On the Conditions of Fatigue in Reading (Griffing and Franz, 1896)[46]

Experiment: As part of several experiments to investigate factors leading to fatigue in reading (including size and quality of type, distance between letters and lines, intensity and quality of illumination) Griffing and Franz used three methods to measure the impact of different colors of paper. The colors used were white, gray-tinted newspaper (white paper with 30 percent black added, yielding a reflectance or relative luminosity of 70 percent), yellow, and red. Each of the colors corresponded to particular line on the color wheel.[47] The number of participants in each experiment was very small, ranging from two (Griffing and Franz themselves) to three.

The first method involved observers viewing a card with three- and four-word phrases on it at a distance of thirty centimeters from their eyes. After being exposed to each card for a period of 1/20 of a second, observers wrote down the words they had seen. The ratio of words seen to the total on each card was then calculated. This method used only white and gray-tinted paper. In the second method, the same apparatus was used but the time it took to "see"[48] all of the phrases, calculating to the thousandth of a second, was recorded. All colors were investigated. The third method measured the illumination necessary to read letters on the cards. A lamp of approximately 0.02 candle power was moved progressively closer to the card to be viewed, which was exposed for ½ seconds between each movement of the lamp. The cards consisted of three lines of ten to twelve words. Only white- and gray-colored papers were used.

Results: For the first method, there was no great difference in percent of words seen on white or gray paper. Of 150 words in 11-point type 32 percent were seen by observers on white paper, and 31 percent on gray paper. In the second method, longer exposure times (in thousandths of a second) were found to be necessary to see all of the phrases on the gray, yellow, and red paper.

---

[46] Griffing, H., and Franz, S. (1896). "On the Condition of Fatigue in Reading"; *Psychological Review*, 3, pp. 513-530. G and F represent the participants, presumably Griffing and Franz.
[47] Griffing and Franz, pp. 528-529.
[48] The definition of this is unclear.

	White.	News.	Yellow.	Red.
G.	2.8	4.0	4.0	—
F.	1.2	1.7	2.5	4.0

Table VIII. Time to Recognize Words on Different Colors of Paper.[49]

In the third method, nearly twice the amount of illumination was needed to view text on the gray paper as the white paper.

Observer.	G		F		F_1		H		H_1	
	Av	MV	Av	MV	Av	MV	Av	MV	Av	MV
W = White	.10	.01	.10	.01	.06	.01	.04	.00	.10	.02
N = News	.20	.02	.16	.02	.08	.01	.07	.00	.23	.01
$\frac{W}{N} = \lambda$	.50		.62		.75		.57		.43	

Table IX. Illumination Thresholds for White and Gray Paper[50]

Conclusions: Griffing and Franz's conclusions for these experiments were:

> If the paper used reflects very little light and is of such a quality that letters can be well printed, the exact hue is probably of little importance, provided a large quantity of light be diffused. But if the absorption be so great that the paper appears grayish [or red or yellow], letters printed on it will not be so legible by reasoning of the lessening of the contrast between the letters and the background.[51]

They also noted in their general conclusions of the study that white paper should be used for best legibility, though it was possible that "the greater amount of light reflected from pure white paper may cause some fatigue."[52]

<u>The Comparative Legibilty of Black and Colored Numbers on Colored and Black</u> Backgrounds (Miyake, Dunlap, Cureton, 1930)[53]

---

[49] Griffing and Franz, p.529.
[50] Griffing and Franz, p.529.
[51] Griffing and Franz, p.528.
[52] Griffing and Franz, p.530.
[53] Miyake, R., Dunlap, J.W., and Cureton, E. E. (1930). "The Comparative Legibility of Black and Colored Numbers on Colored and Black Backgrounds; *The Journal of General Psychology*, 3, pp. 340-343.

Experiment: Two series of text materials were prepared to investigate the effect of colored backgrounds and text on legibility. In the first series, random numerals from one to nine were printed on white, red, green, and yellow slips of paper. Three samples of each color were prepared so there were 12 slips of paper in all. In the second series random numerals from one to nine were colored onto black slips of paper. Again, three samples of each color were made, or 12 slips in all. A spring tachistoscope, a device frequently used in measuring speed of perception,[54] was used to expose the slips of paper in each series to fifteen subjects. The subjects were instructed to write down the numerals they saw on a sheet of paper, guessing if they were not sure, and writing nothing if they did not see anything. A score was calculated based on how many numerals each subject identified correctly on each color of paper (or each color of print, for the second series). The tachistoscope was calibrated so that every subject was able to recognize at least one letter in the time of exposure. Actual exposure time was not measured.

Results: The results are shown below. Subjects had the greatest difficulty seeing black print on a red background in Series I and red print on a black background in Series II.

	Series I				Series II			
Subject	Red	Green	Yellow	White	Red	Green	Yellow	White
A	27	27	27	27	23	20	27	27
B	27	27	27	27	0	0	23	27
C	27	27	27	27	3	10	24	27
D	25	27	27	27	6	8	21	27
E	26	27	27	27	11	1	25	27
F	0	19	25	27	0	1	1	27
G	25	27	27	27	1	0	20	27
H	26	26	27	27	4	3	23	27
I	24	27	27	27	1	7	26	27
J	23	27	27	27	1	3	23	27
K	20	26	26	27	3	1	21	27
L	25	27	27	27	0	1	27	27
M	18	27	27	27	0	0	21	26
N	25	27	27	27	0	0	20	27
O	27	26	27	27	0	0	16	27
Mean	23.00	26.27	26.80	27.00	3.53	3.67	21.20	26.94

Table X. Scores of Subjects by Series and Color[55]

[54] A tachistoscope is an apparatus that allows an "exposure field" to be presented to an observer for a very short period of time (1/10 second or less) and then hidden (Tinker, 1963, p.12). A variety of designs for tachistoscopes exist, including those that make use of projection or a system of mirrors. It is not clear what type of tachistoscope is being used here.
[55] Miyake, Dunlap, and Cureton, p.341.

Miyake, Dunlap, and Cureton presented the significance of the mean differences in these scores in the following table. *d* □ represents the observed mean difference, *t* is a measure of the probability of the significance of the difference (a value the researchers calculated from a table in Fisher, 1925[56]), and *p* is the probability that the difference in means rose by chance. Probabilities less than 0.05 generally indicate a significant difference.[57]

	$\bar{d}$	t	p
	Series I		
$M_G—M_R$	3.27	2.496	.028
$M_Y—M_R$	3.80	2.310	.040
$M_W—M_R$	4.00	2.254	.046
$M_Y—M_G$	.53	1.299	.217*
$M_W—M_G$	.73	1.377	.200*
$M_W—M_Y$	.20	1.355	.190*
	Series II		
$M_G—M_R$	.16	.160	.870*
$M_Y—M_R$	17.69	9.730	.01
$M_W—M_R$	23.40	14.745	.01
$M_Y—M_G$	17.53	10.348	.01
$M_W—M_G$	23.26	16.232	.01
$M_W—M_Y$	5.74	3.517	.01

*Difference not significant.

Table XI. Significance of Mean Differences[58]

As the table indicates, in Series I significant differences in colors were only observed between red and the three other colors (not between black text on yellow vs. green, white vs. green, or white vs. yellow backgrounds). In Series II, all differences with the exception of green text on the black background and red text on the black background were significant. The greatest differences were observed between white and red text, and white and green text.

Conclusions: Miyake, Dunlap, and Cureton concluded that since significant differences were observed in eight out of the twelve test instances, further investigations into color would offer more insight into issues of legibility. They note that the relative illegibility of black letters and a red background and red letters on a black background is clear.[59]

---

[56] Fisher, R. A. (1925). *Statistical Methods for Research Workers*; Oliver & Boyd, London.
[57] Miyake, Dunlap, and Cureton, p.342.
[58] Miyake, Dunlap, and Cureton, p.343.
[59] Miyake, Dunlap, and Cureton, p.343.

Effect of Size of Object and Difference of Coefficient of Reflection as Between Object and Background (Ferree and Rand, 1929)[60]

Experiment: Ferree and Rand investigated two variables, size of object and difference of coefficient of reflection between object and background (reflectance) as part of an experiment to show the effect of changes in light intensity on speed of vision.[61] The objects used in this case were circles, each having an opening at one of eight different positions (up, down, right, left, and each of the four 45 degree positions). The objects had between 3 and 4 percent reflectance and were placed on backgrounds having 78 percent, 29 percent, 21 percent, and 16 percent reflectance, respectively.[62] The sizes of the objects corresponded to visual angles of 1, 2, 3, 4.2, and 5.2 minutes of arc at 2.5 meters, or approximately 21 point, 41 point, 62 point, 87 point, and 108 point. Since the experiment was designed to examine light intensity in particular, great pains were taken to create an environment where factors such as illumination, angle of vision, and subject fatigue could be tightly controlled. Because of the involved nature of the study, the data Ferree and Rand present are those taken from one test case only. This test subject, referred to as “R” (presumably Rand himself), was found to be an average performer in the experiments when compared with other subjects trained in the methods and equipment used in the study.[63]

Using a tachistoscope, subjects were exposed to the objects for short intervals of time at different levels of illumination (ranging from 1.25 to 100 foot candles. At each interval, and for each size and intensity of light, the objects were exposed in each of their 8 positions. Subjects indicated the direction of the opening when an object was exposed, and if a correct judgment was made for 5 out of the 8 positions, they were considered to have been able to discern the object at that time interval.

Results: Ferree and Rand plotted the results of all their experimental trials on one combined graph, and then on separate graphs to ease comparison. The combined graph can be seen in Figure 8 below. The speed of discernment (as a reciprocal of time) is shown on the left side of the vertical axis, the time for

---

[60] Ferree, C. E., and Rand, G. (1929). “Intensity of Light and Speed of Vision, I; *Journal of Experimental Psychology*, 12(5), 363-391.

[61] The speed of the eye’s reactions or “ocular efficiency”. Ferree and Rand, p.381. It should be noted that the experiments by Ferree and Rand were performed primarily for the benefit of the industrial sector (textile manufacture) and not printing, per se.

[62] Ferree and Rand (1929), pp.368-369.

[63] R performed in the upper quartile when compared with “untrained” subjects. The exact difference between trained and untrained subjects is unclear, as it is exactly how many subjects completed the study (the assumption is that several participants were involved, even though the results presented are those of “R” alone). Ferree and Rand, p.367.

discernment on the right side of the vertical axis, and illumination in foot candles on the horizontal axis. The size of the object and percent reflectance are indicated on each line of the graph. Although the graph is difficult to decipher at first glance, it is clear that objects of each size are seen faster on a background of high reflectance (78 percent) than of low reflectance (29, 21, and 16 percent) at each level of illumination.[64]

Ferree and Rand noted this, as well as the fact that “in every case a higher speed is attained with a high coefficient of reflection and a low illumination than with the equivalent brightness of background produced by a low coefficient of reflection and a high illumination.”[65] Their data is shown in Table XII.

[64] Compare the speeds for objects with a size of 3 Min at each reflectance level as an example.
[65] Ferree and Rand (1929), p.383.

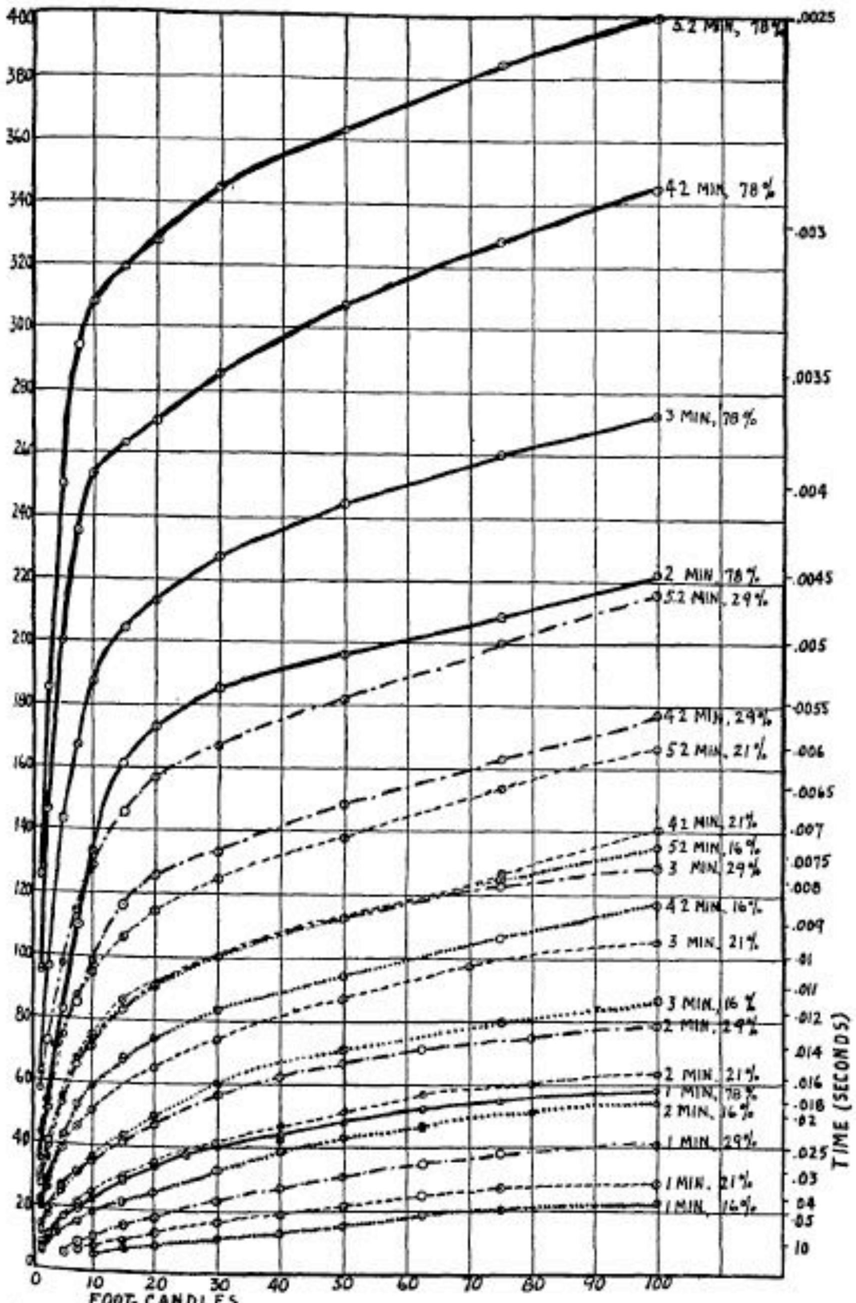


FIG. 1. CURVES SHOWING FOR GROSS COMPARISON ALL THE RESULTS OBTAINED ON THE EFFECT INCREASE OF INTENSITY OF LIGHT ON SPEED OF VISION.

Figure 8. Curves Showing For Gross Comparison All the Results Obtained on the Effect of Increase of Light on Speed of Vision[66]

[66] Ferree and Rand, p.369.

A COMPARISON OF THE SPEEDS OBTAINED WITH EQUAL BRIGHTNESSES OF BACKGROUND, WHICH WERE SECURED BY THE APPROPRIATE COMBINATION OF HIGH COEFFICIENT OF REFLECTION AND LOW ILLUMINATION, OR LOW COEFFICIENT OF REFLECTION AND HIGH ILLUMINATION

Coeff. of refl'n of b'kground (%)	B'ghtness of b'kground (Ml)	Illumination (Ft.C.)	Speed for Visual Angles of 1 Min	2 Min	3 Min	4.2 Min	5.2 Min
78	10	12	25	146	194	259	310
29		32	24	58	102	135	169
21		45	22	50	84	110	135
16		58	18	45	75	100	118
78	17	21	34	173	213	271	328
29		54	32	69	115	152	185
21		76	27	59	98	128	155
16		99	22	54	87	117	135

Table XII. A Comparison of the Speeds Obtained With Equal Brightness of Background[67]

Conclusions: The reasons they cited to explain these results were 1) the effects of differing states of eye adaptation at different levels of illumination, 2) different states of eye adaptation for different pupil sizes, and 3) that "the prime factor in discriminability of the object is not the brightness of the background, but the *difference* in brightness between the object and background."[68]

<u>Intensity of Light and Speed of Vision, II. Comparative Effects for Dark Objects on Light Backgrounds and Light Objects on Dark Backgrounds (</u>Ferree and Rand, 1930)[69]

Experiment: In their previous study, Ferree and Rand described several areas of experimentation that would further their understanding about the effects of light intensity on the eye's speed of response. One of these, an investigation of the effect of light objects on a dark background to compare with results of dark objects on a light background from the previous experiment, was the object of this study. Ferree and Rand prepared white test objects that were exposed on black backgrounds (4 percent reflectance), and gray backgrounds (21 percent reflectance) in the same way, and under the same conditions as the black objects in the experiments performed previously with one exception. It was found that the speed of vision for white objects greater than 3 arc minutes was faster

[67] Ferree and Rand, p.382.
[68] Ferree and Rand (1929), p.383-384.
[69] Ferree, C. E., and Rand, G. (1930). "Intensity of Light and Speed of Vision, II. Comparative Effects for Dark Objects on Light Backgrounds and Light Objects on Dark Backgrounds"; *Journal of Experimental Psychology*, 13, 388-422.

than the equipment was able to measure. For this reason, only objects of 1, 2, and 3 arc minutes were used.

Results: The results are shown in Figure 9 below. Ferree and Rand observed higher speeds of vision, in general, for white objects on black backgrounds than black objects on white backgrounds, and for white objects on gray than for black objects on gray. An exception was that black objects on white were observed faster than white objects on black at low intensities for objects at 1 arc minute. At high intensities, white on black was faster as in other cases.

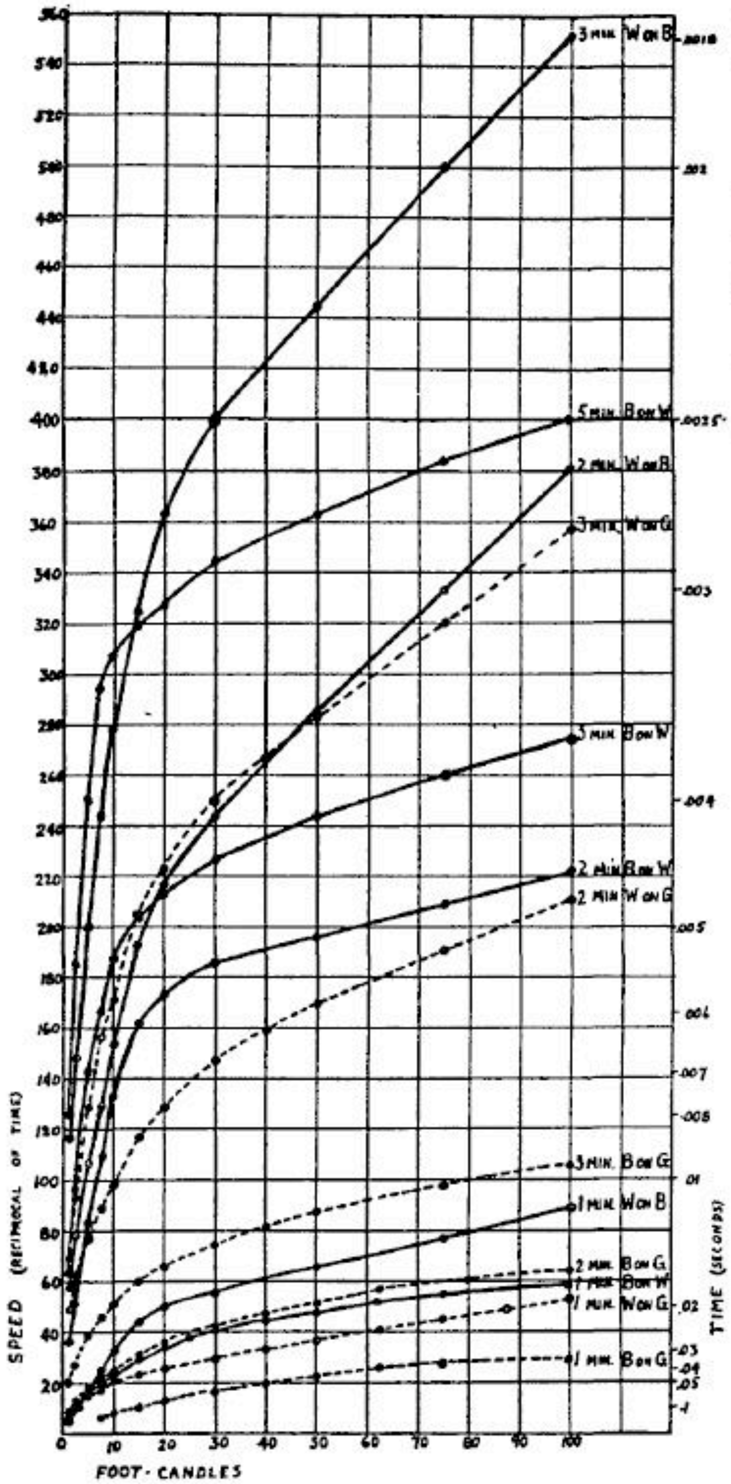


FIG. 1. Curves showing for gross comparison the results obtained on the effect of increase of intensity of illumination on speed of vision for the following cases: 1, 2 and 3 min white test-objects on black background; 1, 2, 3 and 5 min black test-objects on white background; 1, 2 and 3 min white test-objects on gray background; and 1, 2 and 3 min black test-objects on gray background.

Figure 9. Curves Showing For Gross Comparison the Results Obtained On the Effect of Increase of Intensity of Illumination on Speed of Vision[70]

[70] Ferree and Rand (1930), p.393.

Conclusions: Ferree and Rand explained the shorter time needed to discern black objects of 1 arc minute on white at low intensities in terms of visual acuity threshold. Visual acuity was known by test to be higher for black letters on white than white letters on black,[71] and in a situation where visual acuity was paramount (when objects are very small) it would make sense that the black objects would be more easily seen.

The shorter time needed to see white objects at 2 and 3 arc minutes on black backgrounds as illumination increased was explained in terms of sensation differences between white on black and black on white, and the effect of after-images. Ferree and Rand state: “There is a greater difference in sensation between object and background in case of white on black than black on white, due probably to physiological induction or contrast.”[72] Sensation difference is more important than visual acuity at high intensities and for large objects, they held, thus the observed results. Ferree and Rand also determined that after images played a large role. A tachistoscope works by showing subjects a “pre-exposure” field, upon which the exposure-field containing the object to be viewed is projected or reflected. In these experiments, the pre-exposure field for white objects on black backgrounds was black, and the pre-exposure field for black objects on white was white. As Ferree and Rand demonstrate later in the study, the effect of an after image for obscuring the object (given a very short time of exposure) was greater for black objects on white than white objects on black. This helps to account for subjects improved performance with white objects on black at high intensities and large sizes of object.

These same factors (sensation difference and after image) were used to explain the faster discrimination of white objects on gray than black objects on gray. The gray, at 21 percent reflectance, behaved more like the black objects and backgrounds than the white in the experiments described.[73]

---

[71] This is due to the phenomenon of irradiation whereby a white object encroaches on a black background, appearing larger, and a black object is enveloped by a white background, appearing smaller. The result from the standpoint of acuity is that even though black letters would appear smaller on a white background than white on black, the space between and inside of letters would be more defined for the black letters, making them more easily discerned when visual acuity becomes a very important factor, as it does at small sizes of objects or letters. Ferree and Rand (1930), p.394.

[72] This is demonstrated in experiments later in the study, which are not described here. Ferree and Rand (1930), p.397.

[73] Ferree and Rand (1930), p.399.

The Effect of Luminosity On The Apprehension Of Achromatic Stimuli (Taylor and Tinker, 1932)[74]

Experiment: Taylor and Tinker investigated the effect of brightness (reflectance) on the perception of black, dark gray, and light gray letters. A similar methodology to Miyake, Dunlap, and Cureton, described above, was used. Black, dark gray, and light gray consonants measuring 3 by 4 ½ inches were pasted onto two series of white cards, nine to a card. The first series contained 12 cards with letters of homogenous brightness, 4 cards each for black, dark gray, and light gray letters. The second series contained 12 cards with letters of heterogeneous brightness, each card having 3 letters of each brightness in succession (black, dark gray, light gray, etc.). The cards were exposed for three seconds each to a total of 128 university sophomores, who were divided into classes of about 30. After viewing a card, students were asked to write down the letters they could remember and a tally of the total number of letters at each brightness was taken. Equal scores were given for letters reproduced in and out of order.[75]

Results: The mean scores for each series are given below:

Stimulus	Homogeneous		Heterogeneous	
	Mean	σ_M	Mean	σ_M
Black	24.63	.34	25.24	.34
Dark gray	24.70	.31	24.50	.31
Light gray	23.29	.30	21.06	.24

Table XIII. The Influence of Brightness On The Apprehension of Letters
N=128 University Sophomores[76]

Taylor and Tinker found that there was almost no difference between the apprehension of black and dark grey letters in the homogenous series, but there was a significant difference between both black and light gray, and dark gray and light gray. In the heterogeneous series the differences were more clear: apprehension of black was the best, followed by dark gray and then light gray. Calculations of the sizes of the differences in scores, the intercorrelation between scores, and the reliability of the direction of differences revealed that the direction

[74] Taylor, C. D., and Tinker, M. A. (1932). "The Effect of Luminosity on the Apprehension of Achromatic Stimuli"; *The Journal of General Psychology*, 6, 456-458.
[75] Taylor and Tinker found the reliability of the test to be high as a measure of visual apprehension whether misplaced letters were given full or half credit. Taylor and Tinker, (1932), p.457.
[76] Taylor and Tinker, p.457.

of the differences between light gray and dark gray, and light gray and black are correct.[77]

Conclusions: Taylor and Tinker concluded that "there is a direct relation between apprehension scores and luminosity difference between letters and background."[78]

Perceptibility at a distance

Legibility of Colored Print at a Distance (Luckiesh, 1915)[79]

Experiment: In 1915, Matthew Luckiesh reprinted the results of a 1913 study comparing the legibility of different combinations of print and background colors in his book, *Color and Its Applications.*[80] The exact methodology of the study he cited is not clear, but the experiment involved viewing different colors of print on different colors of background at a "considerable distance".[81]

Results: The ranks of the different combinations of print and background are given from best to worst as follows:

1. Black on yellow
2. Green on white
3. Red on white
4. Blue on white
5. White on blue
6. Black on white
7. Yellow on black
8. White on red
9. White on green
10. White on black
11. Red on yellow
12. Green on red
13. Red on green

Conclusions: Information about specific differences between colors is not given so it is difficult to know how much better on color combination was than

[77] Taylor and Tinker, pp.457-458.
[78] Taylor and Tinker, pp.458.
[79] Luckiesh, M. (1915). *Color and its applications*, D. Van Nostrand Co., New York, p.136-137.
[80] Luckiesh, M. (1915).
[81] Luckiesh, M. (1915), p.137. Luckiesh reports the results as having been printed in *Scientific American Supplement*, February 2, 1913. The author was not able to find a February 2 issue of *Scientific American Supplement* (February 1 and February 8 are the closest dates) or locate the actual source of the study. Other researchers, such as F. C. Sumner describes the origins of experiment as being "shrouded in hearsay" (Sumner, 1932, cited below).

another.[82] Luckiesh notes that the "customary" black on white combination is sixth in the list, stating that although the results are interesting, they are not final, "owing to the many variables that enter such a problem."[83]

The Influence of Color On Legibility of Copy (Sumner, 1932)[84]

Experiment: In 1932, a follow-up to the study reported by Luckiesh was performed by F. C. Sumner of Howard University. Sumner expanded the combinations of colors that were used from 13 to 42, and ranked both the legibility of each combination (based on the maximum distance at which six stenciled characters on cardboard backing could be read), and its "affective preference", as determined by the five subjects participating in the study.

Results: The results are shown in Table XIV below, and a comparison of these results with those of the 1913 study is shown in Table XV. It is interesting to note here, as Luckiesh did in the previous study, that many color combinations are ranked higher than black on white.

[82] Tinker notes this short-coming in Tinker, 1963, p.141.
[83] Luckiesh (1915), p.137.
[84] Sumner, F. C. (1932). "Influence of color on legibility of copy"; *Journal of Applied Psychology*. 16(2) pp. 201-204.

CASE NO	BACKGROUND	LETTERING	LEGIBILITY RANK	AFFECTIVE PREFERENCE RANK
1	Gray	Blue	1.0	1.0
2	Gray	Black	2.0	19.0
3	Yellow	Black	3.0	6.0
4	Gray	Red	4.0	3.0
5	Yellow	Gray	5.0	20.0
6	Gray	Green	6.0	9.0
7	Yellow	Blue	7.5	7.0
8	White	Gray	7.5	23.0
9	Gray	Yellow	9.0	14.0
10	Yellow	Red	10.5	10.0
11	White	Black	10.5	16.0
12	Yellow	Green	12.0	18.0
13	Red	Green	13.0	30.5
14	Gray	White	14.0	11.5
15	Green	Blue	15.0	24.0
16	Green	Black	16.5	36.0
17	Green	Gray	16.5	27.0
18	White	Green	18.0	15.0
19	Blue	Green	19.0	25.0
20	Black	Green	21.0	37.0
21	Red	Gray	21.0	26.0
22	Blue	Yellow	21.0	17.0
23	Green	Yellow	23.0	29.0
24	White	Red	24.5	8.0
25	White	Blue	24.5	2.0
26	Red	Black	26.0	42.0
27	Black	Yellow	27.0	5.0
28	Black	Gray	28.0	32.0
29	Blue	White	29.5	13.0
30	Blue	Gray	29.5	11.5
31	Red	Blue	31.0	33.0
32	Red	Yellow	32.0	28.0
33	Blue	Red	33.0	34.5
34	Red	White	34.0	22.0
35	Black	White	35.0	4.0
36	Green	White	36.0	21.0
37	Yellow	White	37.0	34.5
38	Green	Red	38.0	30.5
39	Black	Red	39.0	38.0
40	Blue	Black	40.0	41.0
41	White	Yellow	41.0	40.0
42	Black	Blue	42.0	39.0

Table XIV. Order of Color Combinations by Legibility Rank and Affective Preference[85]

[85] Sumner (1932), p.203.

ORDER OF VALUE (LUCKIESH)	PRINTED MATTER	BACK-GROUND	ORDER-INDEX (PRESENT STUDY)	RANK ORDER (PRESENT STUDY)
1 (greatest)	Black	Yellow	3.0	1.0
2	Green	White	18.0	5.0
3	Red	White	24.5	6.5
4	Blue	White	24.5	6.5
5	White	Blue	29.5	9.0
6	Black	White	10.5	2.5
7	Yellow	Black	27.0	8.0
8	White	Red	34.0	10.0
9	White	Green	36.0	12.0
10	White	Black	35.0	11.0
11	Red	Yellow	10.5	2.5
12	Green	Red	13.0	4.0
13	Red	Green	38.0	13.0

rho .46

Table XV. Comparison of Sumner's results with those reported by Luckiesh in 1915[86]

Conclusions: Sumner came to the following conclusions:

1) The findings of Luckiesh, Poffenberger,[87] and others that legibility depends on the brightness-contrast between printed text and background appeared substantiated.

2) A second law of legibility, that dark colored lettering on a light colored background is more legible than the reverse in daylight, was observed.

3) In his investigations, gray formed the best background for legibility of colored letters.

4) When his results were compared with the experiments reported by Luckiesh, there was a fairly high correspondence (.46) in spite of the fact that exact colors and conditions in that experiment were unknown.

[86] Sumner (1932), p.204.
[87] A. T. Poffenberger was a psychologist who wrote about the attention value of color in advertising in his book, *Psychology in Advertising*; A. W. Shaw Company, Chicago, 1925. In a section entitled "Influence of Colors On Legibility of Copy", he reprints the results of the same experiment Luckiesh did (of 1913) and compares them to the results of a light intensity study done by D. E. Rice to provide a basis for understanding the use of color in advertising. In the course of discussing the use of color combinations to attract attention in advertising, he states, "The general rule can be laid down that legibility depends upon relation of color to background and that the all important factor is *brightness difference*" (Poffenberger, p.263).

5) A number of uncontrollable factors interfered with the investigation of legibility:

1. negative after images observed by subjects
2. irradiation
3. some characters were "misleading"[88]
4. Individual characters varied in legibility
5. A uniform rest interval for clearing effects of after-images and adaptation- and accommodation-effects was difficult to find for all subjects
6. The difference in legibility of some color combinations so slight that they were ranked differently by different subjects
7. A competitive attitude affected the results[89]

6) There was a high positive correlation between legibility and affective preference of color combinations (rho 54).

7) The affective preference of color combinations corresponded more closely to the law of brightness-difference than observed legibility.

8) Affective preference depends more on the brightness difference between text and background than on legibility.

<u>The Effect of Variations in Color of Print and Background on Legibility</u> (Preston, Schwankl, and Tinker, 1932)[90]

Experiment: Preston, Schwankl, and Tinker investigated the furthest distance from the eye that five-letter words printed in colored ink on different colors of paper could be read accurately. 11 color combinations were used. The trade names of the papers and inks, as well as the observed effects of combining them, are shown below.[91]

---

[88] It is not clear what Sumner means in this statement (Sumner, 1932, p.202).
[89] Sumner does not say how, exactly, but it can be imagined that those who "tried" harder were able to see further.
[90] Preston, K., Schwankl, H. P., and Tinker, M. A. (1932). "The Effect of Variations in Color of Print and Background on Legibility"; *The Journal of General Psychology*, 6, pp. 459-461.
[91] Preston, Schwankl, and Tinker (1932), p.459.

Trade name	*Observed effect*
Black jobbing on white	(Black on light grayish white)
Grass green on white	(Dark green on light grayish white)
Lustre blue on white	(Dark blue on light grayish white)
Black jobbing on yellow	(Black on yellow with slight orange tinge)
Tulip red on yellow	(Light red on yellow with slight orange tinge)
Tulip red on white	(Light red on light grayish white)
Grass green on red	(Dark grayish green on red)
Chromium orange on black	(Dark lemon yellow on dark grayish black)
Chromium orange on white	(Light orange on light grayish white)
Tulip red on green	(Dark brown on dark green)
Black jobbing on purple	(Black on dark purple)

For each combination, 4 lines of 4 words each were printed in random order on 4 sheets of paper. 66 study participants were each given one sheet containing a color combination and one sheet of black text on the white background (the black on white sheet served as a standard of comparison). 6 subjects, then, compared each color combination with black on white. To conduct the study, the sheets were placed in a carriage mechanism at the end of a long bench and moved progressively closer to each participant at 20 cm intervals until every word on the sheet could be read accurately. Care was taken to control for light intensity, fatigue, and practice effects.

Results: Preston, Schwankl, and Tinker found the differences between the color combinations to be as follows (minus differences indicate that the average distance at which words were correctly read for the color combinations was greater than for black text on white, indicating greater legibility, plus differences indicate the opposite):

Blue on white: all minus differences
Black on yellow: all minus differences
Green on white: 4 minus, 2 plus differences
Green on red: 1 minus, 5 plus differences
Red on yellow: 1 minus, 5 plus differences
Red on white: all plus differences
Orange and black: all plus differences
Black on purple: all plus differences
Orange on white: all plus differences
Red on Green: all plus differences
Red on white: all plus differences[92]

[92] Preston, Schwankl and Tinker (1932), p.460.

Data for the experiments is shown in the table below:

Color of ink and paper	Legibility rank	Color combination Mean	σ_M	Black on white Mean	σ_M	Diff.	$\frac{Diff.}{\sigma_{diff.}}$
Blue on white	1	162.4	.22	150.6	.20	—11.8	3.97
Black on yellow	2	166.5	.27	155.7	.25	—10.8	2.96
Green on white	3	182.5	.24	177.1	.22	— 5.4	1.66
Black on white	4	166.6*	.22	166.6*	.22	0.0	0.00
Green on red	5	164.0	.20	169.7	.22	+ 5.7	0.57
Red on yellow	6	140.6	.19	150.2	.18	+ 9.6	3.66
Red on white	7	161.1	.24	179.9	.25	+18.8	5.42
Orange on black	8	152.7	.19	172.7	.19	+20.0	7.49
Black on purple	9	99.7	.14	161.1	.22	+61.4	23.52
Orange on white	10	91.8	.12	178.2	.18	+86.4	40.00
Red on green	11	78.3	.14	170.4	.23	+92.1	34.36

*This mean for black on white is an average of means for the ten subgroups.

Table XVI. The Effect of Variations In Color Combinations On The Legibility Of Print [The mean score is the average distance in centimeters from the eye at which the words were read [each mean is an average of 384 scores][93]

The next to last column shows the difference between the means for color combinations and the means for black on white. The last column gives the ratio of each difference to its standard error.

Conclusions: Preston, Schwankl and Tinker observed that the color combinations that ranked highest were those with the greatest brightness contrast between print and background. They concluded that "the greater the luminosity or brightness differences between symbol and background, the greater the legibility of the print."[94]

Speed of Reading

<u>Studies of Typographical Factors Influencing Speed of Reading VII. Variations In Color Of Print And Background</u> (Tinker and Paterson, 1931)[95]

Experiment: Tinker and Paterson used the Chapman-Cook Speed of Reading test[96] to measure the effect of print and background colors on legibility.

[93] Preston, Schwankl and Tinker (1932), p.460.
[94] Preston, Schwankl and Tinker (1932), p.461.
[95] Tinker, M. A., and Paterson, D. G., Studies of Typographical Factors Influencing Speed of Reading. VII. Variation in Color of Print and Background. *Journal of Applied Psychology*. 1931, 15, 471-479.

This test consists of 30 paragraphs of 30 words each printed in two columns on an 8 ½ by 11 inch sheet of paper. In the second half of every paragraph is a word that somehow "spoils the meaning" of each paragraph.[97] Subjects are asked to cross out this word when reading the test as a check on reading comprehension, and the number of paragraphs. The spoiler words are chosen so that it is not possible to know which word to cross out without reading the entire paragraph.[98]

Two forms (Form A and Form B) of the test are prepared, with the variables to be measured, such as font size or type, differing from one form to the other. In Tinker and Paterson's experiment, Form A consisted of black ink printed on white Rainbow coverstock[99] (the standard of comparison in the study). Form B consisted of Ruxton's colored ink on Rainbow coverstock in the combinations below:

---

[96] This test, along with a defense against possible short-comings, is discussed in detail in Tinker, M. A. and Paterson, D. G. (1938). "Studies of Typographical Factors Influencing Speed of Reading XIII. Methodological Considerations"; *Journal of Applied Psychology*, 20,1, pp.132-145.
[97] Tinker, M. A. and Paterson, D. G. (1928). "Influence of Type Form On Speed Of Reading"; *Journal of Applied Psychology,* 12,4, p.360.
[98] Tinker and Paterson (1928), p.360. It is the use of this test, which controls for reading comprehension, that sets Tinker and others apart from those who take legibility to refer to the appearance of characters only. The Chapman-Cook test and has been criticized, however, for not requiring a high enough level of comprehension (Pearson, P. David, Barr, Rebecca, Kamil, Michael L., Mosenthal, Peter. *Handbook of Reading Research (vol.1);* Lawrence Erlbaum Associates, 2002, p.24. The authors mention criticism, but do not give a source.)
[99] It is believed that Rainbow is a particular brand of paper.

Column 1 shows test group, column 2 trade names for Ruxton's ink and Rainbow cover-stock used, and column 3 the observed effects of colored ink and paper combinations

Test Group	Trade Names	Observed Effect
(1)	(2)	(3)
I.	Grass green on white	Dark green on light grayish white
II.	Lustre blue on white	Dark blue on light grayish white
III	Black jobbing on yellow	Black on yellow (slight orange tinge)
IV.	Tulip red on yellow	Light red on yellow (slight orange tinge)
V.	Tulip red on white	Light red on light grayish white
VI.	Grass green on red	Dark grayish green on red (dark tint)
VII.	Chromium orange on black	Dark lemon yellow on dark grayish black
VIII.	Chromium orange on white	Light orange on light grayish white
IX.	Tulip red on green	Dark brown on dark green
X	Black jobbing on purple	Black on dark purple (violet)

Table XVII. Test Groups, Color Combinations Of Ink And Paper, And Observed Color Effects[100]

10-point, Scotch Roman type was used. There were 850 people in the study in all, split into 10 groups of 85. Each person completed 4 forms in the order A B B A and the average number of paragraphs read in 1 ¾ minutes was recorded.

Results: The results are presented in Tables XVIII and XIX.

[100] Tinker and Paterson (1931), p. 473.

Influence of Different Combinations of Colored Print and Colored Paper

Differences given are the mean score on Form A using black print on white Rainbow cover-stock, minus the mean score on Form B varying with respect to color of ink and Rainbow cover-stock used. All test blanks were printed in Scotch Roman, 10 point type, 19 pica line width, set solid. The mean score is the average number of paragraphs of 30 words each in the Chapman-Cook Speed of Reading Tests read in 1¾ minutes. In each sub-group, N = 85 college students.

Test Group	Test Form, Color of Ink and Cover-stock	Mean	PE_{dist}	$P.E_M$	Difference between means in: Paragraphs	Words per Sec.	Per Cent	PE^*_{diff}	$\frac{D}{P.E._{diff}}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
I	A, Black on White B, Green on White	17.26 16.75	2.94 2.58	.32 .29	0.51	0 15	3.0	.16	3.24
II	A, Black on White B, Blue on White	16.61 16.06	2.65 2.39	.29 .26	0.55	0.16	3.4	.17	3.24
III	A, Black on White B, Black on Yellow	16 37 15.77	2.90 2.42	.31 .26	0.60	0.17	3.8	.15	3.89
IV	A, Black on White B, Red on Yellow	17 09 16.31	2.47 2.25	.27 .24	0 79	0.23	4 8	.17	4.74
V	A, Black on White B, Red on White	17 65 16.20	3.41 2.80	.37 .30	1.45	0.41	8.9	.20	7.19
VI	A, Black on White B, Green on Red	17.91 16.20	2.70 2.39	.29 .26	1.71	0.49	10.6	.18	9 69

Test Group	Test Form, Color of Ink and Cover-stock	Mean	PE_{dist}	PE_M	Difference between means in: Paragraphs	Words per Sec.	Per Cent	PE^*_{diff}	$\frac{D}{PE_{diff}}$
VII	A, Black on White B, Orange onBlack	16.94 14.93	2.78 2.15	.30 .23	2.01	0.57	13 5	.17	12.05
VIII	A, Black on White B, Orange on White	16 62 13 75	2.60 2.53	.28 .27	2.87	0.82	20.9	.19	15.46
IX	A, Black on White B, Red on Green	17 38 12.46	2.50 2.82	.27 .31	4.92	1.41	39.5	.25	19.91
X	A, Black on White B, Black on Purple	16 66 11.00	2 52 3.08	27 .33	5 66	1 62	51.5	.28	20.29

* $P.E._{diff}$ values in column 9 were computed on the basis of the standard formula for determining the statistical significance of obtained differences for *correlated measures*. The correlations for the ten sub-groups in this experiment were +0.85 ± .02 for Group I; +0.81 ± .03 for Group II; +0.87 ± .02 for Group III; +0 79 ± .03 for Group IV; +0.84 ± .02 for Group V; +0 80 ± 03 for Group VI; +0.83 ± .02 for Group VII; +0.78± .03 for Group VIII; +0.64 ± .04 for Group IX; and +0 59 ± .05 for Group X.

Table XVIII. Influence of Different Combinations of Colored Print and Colored Paper[101]

In each group, Tinker and Paterson found that more paragraphs were read in average for black on white in comparison with the other color combinations. They noted that the differences between average paragraphs read in the first three groups were “not statistically certain”, meaning that green on white, blue on white, and black and yellow might be found to be better in some cases (one or

[101] Tinker and Paterson (1931) p.474. These results were presented in two tables in the original article. The correlation referred to in the caption refers to the correlation of Form A to Form B in each case. It is unclear how this correlation was determined.

two in a hundred) than black on white if the experiment were to be repeated.[102] However, the differences increase for groups IV through X.

Tinker and Paterson compared their results to those of the 1913 reported by Luckiesh (above) and were surprised, as Sumner was a year later, to find a significant amount of agreement given the large number of variables that were unaccounted for in the previous study (font size, type, exact colors used, etc.).

Color Combinations	RANK ORDER ACCORDING TO	
	Present Study	Luckiesh
Black on white	1	5
Green on white	2	2
Blue on white	3	4
Black on yellow	4	1
Red on yellow	5	6
Red on white	6	3
Green on red	7	7
Red on green	8	8

Table XIX. Comparison of Tinker/Paterson results with those cited by Luckiesh[103]

Conclusions: Tinker and Paterson found their results to be in agreement with the conclusion psychologist A. T. Poffenberger (see footnote 87 above) had come to regarding legibility and color. Pointing to the fact that the ranks of the result groups in the experiment had been arranged according to brightness difference, they concluded: “The evidence in this experiment justifies the following rule: In combining colors (color of ink and paper) care must be taken to produce a *printed page* which shows a maximum *brightness contrast* between print and background.” Based on this, they provided a “rough guide” for advertisers:

> Providing good legibility. Black on white, grass green on white, luster blue on white, and black on yellow.
> Providing fair legibility. Tulip red on yellow, tulip red on white.
> Providing poor legibility. Grass green on red, chromium orange on black, chromium orange on white, tulip red on green, black on purple.[104]

[102] Tinker and Paterson (1931), p. 476.
[103] Tinker and Paterson (1931), p. 477.
[104] Tinker and Paterson (1931), p.479.

In a later publication,[105] Tinker pinpoints more specifically the relation between brightness contrast and legibility. In speaking about the relative legibility of print on different colors, he says,

> Dark colored inks coordinated with colored tints of paper can be as legible as black print on white paper provided (a) the reflectance of the paper is 70 per cent or greater, (b) the colored ink has a reflectance low enough so that the brightness contrast between print and paper is about 65 per cent (i.e., 1 to 8 ratio), and (c) the size of type is 10 point or larger.

It is unclear which experimental results Tinker bases this statement on, but it is likely that it comes from the results of the experiment just described. In Tinker (1963), Tinker says that differences of 2 to 5 percent in the number of paragraphs read for this experiment were not statistically significant. Paterson and Tinker did not give the brightness contrasts for the colors of paper that they used, but the 1 to 8 ratio Tinker designates probably comes from the observed difference in reading speed for red on yellow and red on white (or green on red) in the above results.

Further Study: Additional experiments were undertaken in connection with these results. Tinker investigated the judged legibility of the same print and background samples by asking 210 readers their opinions of the relative legibility of each sample (Table XVI). He found a close correspondence between these opinions and his previous results and concluded:

> It would seem that readers make their judgments of relative legibility largely in terms of brightness contrast between print and paper without being influenced by color preference, and color contrast. From a practical point of view, the editor will choose colors which produce maximum brightness contrast when combined, if he is to achieve good legibility and reader approval.[106]

[105] Tinker (1963), p.150.
[106] Tinker (1963), p.148. Neither the date of this study nor the exact investigators are given.

Color Combination	Average Rank	Rank Order
Black on white	2.1	1 (best)
Blue on white	2.8	2
Black on yellow	2.9	3
Green on white	4.2	4
Red on yellow	5.3	5
Red on white	5.4	6
Green on red	5.7	7
Orange on black	7.6	8
Orange on white	9.1	9
Black on purple	10.2	10
Red on green	10.5	11

Table XX. Judgments of Relative Legibility of Colored Print on Colored Paper

In 1944, Tinker and Paterson returned to the results of their 1931 study to investigate, using a measure of eye movements, the reason that red print on green background had been read 39.5 percent slower than black on white (Table XI).[107] They found significant differences in pause duration, perception time, and regression frequency between the two color combinations (up to 28 percent greater for red on green).[108] Hackman and Tinker performed more in-depth investigations of eye movements using nearly all of the color combinations of the first study and recorded the following overall rankings of color combinations:[109]

Color Combination	Mean Rank	Final Rank
Black on yellow	1.75	1 (best)
Red on white	2.00	2
Green on red	3.00	3
Black on white	3.25	4
Black on purple	5.00	5
Orange on white	6.00	6
Red on Green	7.00	7

Table XXI. Ranking of Color Combinations According to Differences Observed in Fixation Frequency, Pause Duration, and Perception Time

[107] Tinker, M. A. and Paterson, D. G. (1944). "Eye Movements In Reading Black Print on White Background and Red Print on Dark Green Background"; *American Journal of Psychology*, 57, pp.93-94. Cited in Tinker (1963, p.148).
[108] Tinker, 1963, p.149.
[109] Hackman, R. B. and Tinker, M. A. (1957). "Effect of Variations in Color of Print and Background Upon Eye Movements in Reading"; *American Journal of Optometry and Archives of the American Academy of Optometry*, 34, pp.354-359. Cited in Tinker (1963), p.149.

Hackman and Tinker found that these rankings correspond closely with the results from 1931, but that in general the measure of eye movements did not show as precise distinctions between the color differences as the measure of reading speed. For this reason they concluded that the investigation of eye-movements with regard to print and background color was a useful supplement to speed of reading tests, but not a valid replacement.[110]

The Visibility of Print on Various Qualities of Paper (Luckiesh and Moss, 1938)[111]

Experiment: Luckiesh set out to measure the legibility (what he referred to as readability)[112] of black print on different colors of paper using three methods: the Luckiesh-Moss visibility meter,[113] speed of reading, and counting the blinks of the eye when reading.[114] The papers that were used and their respective reflectance values are given in Table XV. Four of these, papers *A, I, F,* and *J*, were used in the speed of reading test.

---

[110] Tinker (1963), p.150.

[111] Luckiesh, M., and Moss, F. K. (1938). "Visibility and Readability of Print on White and Tinted Papers"; *Sight-Saving Review*, 8, pp. 123-134. Summarized in Tinker, 1963.

[112] Luckiesh and Moss (1938), p.124.

[113] As described above, this consists of two colorless photographic filters that can be rotated in front of the eyes to reduce both the brightness of the visual field and lower the brightness contrast between an object and its background. A score of "1" on the visibility meter corresponds to the point where a detail of 1 arc minute is visible to an observer of "normal vision" at an illumination of 10 foot candles. A score of "2" corresponds to the point where the test object is visible when it subtends a visual angle of 2 minutes, and so forth.

[114] Although Tinker (1963, pp. 17-19) and Zachrisson (1965, p.60) did not consider eye blinks as a valid method of measuring legibility, at the time this paper was written, Luckiesh and Moss maintained that it was an appropriate and highly sensitive criterion upon which to base determinations of legibility.

Paper	*Reflection-Factor (In Per Cent)*	*Trade Name*
A	85	A & G Special Finish—White (No appreciable tint)
B	70	Blue American Eggshell Text Wove (70 lb.) (Light blue-green tint)
C	82	White North Star Dull Coated (Oxford, 80 lb.) (Slight sepia-cream tint)
D	71	Melon American Eggshell Text Wove (70 lb.) (Reddish buff verging on salmon)
E	81	White Rumford Enamel (Oxford, 80 lb.) (Very slight sepia tint)
F	79	Cochin Yellow D & C Torchon (70 lb.) (Fairly saturated yellow)
G	83	White Oxford Antique Book (60 lb.) (Slight cream tint)
H	74	Green American Eggshell Text Wove (70 lb.) (Light yellowish green tint)
I	82	India Oxford Antique Book (60 lb.) (Deep cream tint or very light buff)
J	38	Cambo Red D & C Torchon (70 lb.) (Fairly saturated yellowish red or reddish orange)

Table XXII. Types of Paper With Reflectance Values[115]

Results: Luckiesh observed that the maximum difference between the white papers *A*, *C*, and *E* was 0.24, or 6 percent (see Table XXIII). This is a statistically significant difference, indicating that different textures, weights, and finishes of these white papers do have an effect on the visibility of print. This is especially seen in the case of paper *J*, whose visibility (and reflectance) are much lower than paper *A*. In comparing the two papers, Luckiesh and Moss estimated that viewing 10-point on paper *J* would be equivalent to viewing 6-point type on paper *A*. The differences are always what one would expect, however. Paper *I*, though it has a lower reflectance that paper *A*, recorded a greater visibility; paper *G* and paper *I*, although *G* has a higher reflectance, scored equally in visibility.

[115] Luckiesh and Moss (1938), p.128.

Subject	Relative Visibility									
	A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
1....	5.04	4.52	5.10	4.76	4.72	4.94	5.29	4.98	5.34	2.80
2....	4.67	4.18	4.64	4.37	4.44	4.65	4.71	4.13	4.70	2.80
3....	2.87	2.73	2.90	2.67	2.78	2.91	2.92	2.82	2.99	1.91
4....	4.69	4.31	4.84	4.61	4.58	4.82	5.16	4.76	5.04	2.87
5....	5.77	5.21	5.62	5.27	5.38	5.87	6.07	5.48	6.10	3.34
Average	4.61	4.19	4.62	4.34	4.38	4.64	4.83	4.43	4.83	2.74
Per cent Probable Error	1.2	1.2	1.3	1.1	1.1	1.1	1.0	1.2	1.2	.9

Table XXIII. The Relative Visibility of 10-point Linotype Textype Printed Upon Ten Tinted Papers, Including White and Near White Papers

The results of the speed of reading test, in which words per minute were recorded for reading samples on 4 papers are shown in Table XXIV. After taking the test, subjects were asked to give their opinions about the papers. All preferred white over yellow, and “emphatically” disliked red.

Luckiesh and Moss noted several things: 1) the differences in speed of reading between the 4 papers varied only on the order of 5 percent, while the differences in reflectance from papers *A* to *J* varied by 40 percent, 2) black print on yellow paper was read at a slower rate than black print on white even though the two papers had approximately the same visibility measure, 3) In spite of the extreme dislike subjects had for the red paper as compared to yellow, the speeds of reading on each paper were very close.

	A (*white*)	*I* (*cream*)	*F* (*yellow*)	*J* (*red*)
Words per minute................	284	273	266	268
Relative words per minute.........	100.0	96.2	93.6	94.4
Per cent probable error...........	..	1.0	1.3	1.4

Table XXIV. Words Per Minute Recorded for 20 Subjects Reading For 5 Minutes Each On Two Separate Occasions[116]

Conclusions: The only concrete conclusion stated by Luckiesh and Moss was that the red paper performed the worst of all of the papers in all experiments. They used their results in general, however, to argue against the speed of

[116] Luckiesh and Moss (1938), p.131.

reading test as a valid test of legibility. They cited in particular the nearly equal reading times for the yellow and red papers, and the closeness of all times for reading (5 percent difference) despite the large difference in visibility (40 percent). Unfortunately, no indication is given by Luckiesh and Moss that they controlled for comprehension in their tests of reading speed. In addition, although tests for reading were performed on 20 subjects, this is far fewer than the hundreds consulted by Paterson and Tinker. Both of these elements (controls for reading comprehension and a sufficiently large pool of subjects) are pitfalls that Tinker states must be avoided for tests of reading speed to be valid in measuring legibility, and reasons why the findings of Luckiesh and Moss have been dismissed by others.[117]

Further work: Additional studies were done by Patterson and Tinker (1936)[118] and Luckeish and Moss (1941)[119] on small variations in paper color and quality. Paterson and Tinker investigated the effect of paper surface on legibility (dull versus glossy finish) and found no significant differences in reading performance for white or slightly tinted yellow papers with different finishes. Luckiesh and Moss examined the effects of different tints of "white" paper on legibility and came to the same conclusion: "…degrees of visibility obtained with various grades and finishes of so-called "white" papers are not radically different when the quality of the paper is optimum in each case."[120]

Both of these studies pointed out that their findings ran contrary to opinions current at the time about paper tint and surface, i.e., that less glossy paper or slight yellowish tints were better for reading.

Research Conclusions

Although investigations into the effects of print and background color on legibility have taken different forms and been undertaken on the basis of a variety of criteria, the experimental evidence clearly demonstrates that legibility of print is directly related to the difference in brightness contrast (not necessarily color) between text and background. Tinker has placed the threshold for these

---

[117] Tinker (1963), p.22.

[118] Patterson, D. G. and Tinker, M. A. (1936). "Studies of Typographical Factors Influencing Speed of Reading: XII. Printing Surface"; *Journal of Applied Psychology*, 20, pp. 128-131. Stanton and Burrt (1935) also did work on printing surfaces, reaching the same conclusions as Paterson and Tinker with regard to white and. yellow-tinted paper (Stanton, F. N., and Burtt, H. E. (1935). "The Influence of Surface and Tint of Paper on Speed of Reading"; *Journal of Applied Psychology*, 19, pp. 683-693).

[119] Luckiesh, M. and Moss, F. K. (1941): "The Visibility of Print on Various Qualities of Paper." Journal of Applied Psychology, 25(2), pp.152-158.

[120] Luckiesh and Moss (1941), p.157.

differences in reflection to be about 65 percent, or a ratio of 1:8, before a decline in reading performance is observed. Another component, just as important to the end-product of printing, if not to legibility itself, has also been observed however. This is the judgment by the reader of aesthetic value.

In his studies of legibility through the distance method, Sumner (1932) found that "there was a high positive correlation between legibility and affective preference of color combination," but that the "affective preference of color combinations corresponded more closely to the law of brightness-difference than observed legibility." Tinker and Paterson (1931) found close correlation between readers' opinions about legibility and measured results also:

> It would seem that readers make their judgments of relative legibility largely in terms of brightness contrast between print and paper without being influenced by color preference, and color contrast.[121]

There was not perfect agreement in the results of either study, however, and experiments by Tinker and Paterson in other areas of typography, particularly in speed of reading with different typefaces, have shown that there can be significant differences between reader opinions and measured results.[122] This observation caused Tinker and Paterson to warn against the use of reader preferences as determinants of legibility, asserting that "…mere opinions concerning matters of typography are unsafe guidelines."[123] They allowed, on the other hand, that there was a "practical value" to reader opinions "that should not be overlooked by the printer who desires to cater to the preferences of his readers."[124] This practical value has to do with the fact that differences between reader preferences and measured legibility, though significant, are sometimes not all that large; when the two are at odds, printers looking to sell copies of their materials may well choose to defer to reader preferences.

Tinker and Paterson's position on this was more clearly stated following a comprehensive series of experiments investigating the relationship between actual legibility, judged legibility and "pleasingness". They found a very high correlation between how study participants judged the legibility of various aspects of text presentation (including combinations of colored print and colored

---

[121] Tinker (1963), p.148. Neither the date of this study nor the exact investigators are given.
[122] Tinker, M. A. and Paterson, D. G. (1942). "Reader Preferences and Typography"; *Journal of Applied Psychology*, 26(1), pp. 38-40.
[123] Paterson, D. G. and Tinker, M. A. (1940). *How To Make Type Readable*; Harper & Brothers Publishers, New York, p.19.
[124] Paterson and Tinker (1940), p.19.

paper) and how they rated their "pleasingness".[125] These ratings differed from actual measured legibility, but led them to conclude the following:

> The results presented above [the high correlation between judged legibility and pleasingness] provide a definite answer to those inclined to believe that aesthetic values should have greater weight than "efficiency" in determining printing specifications. The printer should be guided by the facts regarding the speed with which particular typographical arrangements can be read, and also by reader judgments of legibility. When a printing arrangement is shown to promote rapid reading and readers judge this arrangement to be legible, the printer, presumably would employ it. When two or more printing arrangements are equally legible, the printer presumably would employ the one judged to be most legible. However, when the most efficient printing arrangement is judged to be less legible than another, then the printer will be forced to decide whether or not he will cater to the opinions of the readers. In any event the printer's problem is simplified by the fact that readers place high aesthetic values on those printing arrangements which appear to be the most legible.[126]

Conclusion

Where do all of these experiments and findings leave us in relation to the desire to capture grayscale or color versions of print in large-scale digitization? If one were to look at the results alone, the conclusion is quite easy: make sure there is an adequate difference in brightness contrast in the scans that are taken, and legibility requirements for producing print copies from those scans will be met. The same conclusion would apply to digital representations of materials to be displayed on computer screens. When one looks at the processes by which these conclusions were reached, however, and the place of studies relating to foreground and background contrast in the context of legibility as a whole, a slightly more complicated picture emerges.

These studies represent work in only one of the 18 categories described by Pyke that make up the legibility field. Their results are substantiated by at least 5 different, yet valid, methods of investigation, none of which replicates a typical reader experience. Moreover the issue of how reader preferences are to

[125] Tinker, M. A. and Paterson, D. G. (1942). "Reader Preferences and Typography"; *Journal of Applied Psychology*, 26(1), p.38. No attempt was made in their studies to define the word "pleasingness".
[126] Paterson and Tinker (1942), p.40.

be negotiated and/or incorporated into print design remains a large question mark. Although it is not presented here, legibility research on computer screens has met with similar challenges. Certain elements affecting legibility can be isolated and measured, such as brightness or luminance contrast, type face, line spacing, etc., but experimental conditions have yet to harness and control a number of variables that are of increasing importance in the present and future. These include changing reading practices (for printed materials versus computer screens, and for different generations of users), purposes for reading (leisure versus work, detailed analysis versus skimming) and types of materials (textbooks versus newspapers, government records versus graphic novels).

Research exposes these variables at the same time that it produces concrete results for legibility, and our practices for digitization should be an exact reflection: they should be founded in research so that they meet a baseline of accessibility for variables that we do know, and be designed with the flexibility and extensibility to accommodate those we do not. Websites today have links that allow users to toggle the level of contrast at which site content is viewed; large-scale digitization partners offer content in PDF, page-image, and plain text formats. What will the future of reading look like when 10 million volumes from the world's greatest research institutions share the same space as Facebook, Yahoo! Answers, and Wikipedia? What needs will there be? What preferences? We do not know the answers to these questions, but a growing body of research is providing guidelines and minimum standards that we know must be met to ensure access to the widest possible audience whatever the future may hold.

The use of grayscale or color capture for print content has the potential to produce excellent results for legibility. But in order for it to do so we need to give more attention to the brightness contrast of the scans we produce. Sample scans from the Open Content Alliance presented above illustrate this clearly. They were taken from the three most downloaded books as of the date of access, and all had issues of accessibility at one level or another, as did their black and white alternatives. We cannot expect that legibility of printed versions of these volumes (whether printed from the website in PDF form or ordered by patrons as print-on-demand books) would be improved in any way.

Technological and economic constraints will always be a factor in the decisions we make about preserving and accessing collections, but we must be sure in each step as we go forward that the results we produce meet minimum standards of accessibility for our users. Once this minimum is reached, we will be free to imagine and create new modes of access to satisfy the increasingly diverse needs of readers and researchers in the 21^{st} century.