1-1-2016

E-Book Reading Practices in Different Subject Areas: An Exploratory Log Analysis

Robert S. Freeman
Purdue University, rsfree@purdue.edu

E. Stewart Saunders
Purdue University, ssaunder@purdue.edu

E-Book Reading Practices in Different Subject Areas: An Exploratory Log Analysis

Robert S. Freeman

E. Stewart Saunders

Abstract

Print books pose inherent difficulties for researchers who want to observe users' natural in-book reading patterns. With e-books and logs of their use it is now possible to track several aspects of users' interactions inside e-books, including the number and duration of their sessions with an e-book and the order in which pages are viewed. This chapter reports on a study of one-year of EBL user log data from Purdue University to identify different reading patterns or ways in which users navigate within different types of e-books—authored monographs vs. edited collections--and in e-books in different subject areas. The analysis of reading logs for e-books is still very much a new venture. From this perspective the results of this chapter are exploratory and descriptive, rather than conclusive, and as much about the evolution of workable methodologies as they are about the results of the analysis. Log analysis reveals nothing about users' circumstances or intentions; however, in tandem with usability studies, and studies based on surveys, diaries, and interviews, it can contribute to a more objective understanding of users' interactions with e-books.

Background and Introduction

In the ancient world reading was usually done out loud. In *A History of Reading*, Alberto Manguel (1996) recounts a story from the *Confessions* of St. Augustine where

Augustine tells of the time he paid a visit to Ambrose, the Bishop of Milan. Augustine observed Ambrose reading: “his eyes scanned the page and his heart sought out the meaning, but his voice was silent, and his tongue was still” (*Confessions* 6, 3, as cited by Manguel, 1996, p. 42). This was remarkable to Augustine because reading silently was something out of the ordinary.

Like Augustine’s observation, most objective descriptions of silent reading have focused on its physiognomic aspects, (i.e., reading posture, facial expression, movement of the hands, fingers, tongue, lips, and eyes). In the 19th and 20th centuries, many scientific studies of reading concentrated on readers’ visual behavior or eye-movements. Methods of tracking eye movements included the corneal reflection and the scleral observation methods, both of which required holding the subject’s head in a fixed position. Other methods involved attaching monitors to the subject’s eye while the subject scanned a page or read lines of text. Another study placed the reader in a darkened room with a text and a flashlight. “The use of a light is clearly somewhat unnatural for the reader,” the educational psychologist A. K. Pugh (1977) noted, “but the restrictions on the subject are less than in most of the eye-movement recording methods” (p. 42). Pugh discussed a fundamental discovery resulting from Louis-Émile Javal’s early eye-movement studies; when reading or scanning, human eyes do not move smoothly, but rather make jerky movements (saccades) and stop several times, moving very quickly between each stop (fixation). The movements measured in these experiments are very small, and the subjects read only relatively short texts (Pugh, 1978, p. 14). Marshall (2009) notes that, although eye-tracking “provides important data

about some aspects of reading—word and letter recognition, most importantly—it has not shed as much light on how people read in the wild," that is, read naturally (p. 101).

Other controlled reading studies give test subjects identical reading material with instructions, observe and record their actions (i.e. through video recording), and, in some studies, ask them all the same series of questions. User studies often are conducted to inform improvements in the design of products, including printed and digital documents and web pages. A study by Liesaputra and Witten (2008) compared users' interactions with print books and different e-book formats, including one that simulated a 3-D book with realistic page turning. Still, the nature of silent reading makes it difficult to study and measure in the laboratory. The fact that the act of observing affects the behavior being observed means that such research can only go a short way towards describing reader behavior. Reading researchers have long recognized the need for observations or field work in natural situations.

In *Reading and Writing the Electronic Book*, C.C. Marshall (2009), who has observed natural reader behaviors for Microsoft Research, identifies the following kinds of field studies: surveys and questionnaires, interview and diary studies, and studies using instrumenting software that logs details of user interactions with digital technologies such as e-books.

Since the advent of e-books, academic librarians have been conducting surveys to determine how well e-books are catching on with students and faculty. Among the larger surveys of students and faculty by librarians are Levine-Clark (2006), who received 2,067 responses at the University of Denver; Nicholas and colleagues (2008), who received 1,818 responses at University College London; Li, Poe, Potter, Quigley

and Wilson (2011), who received 2,569 responses from the University of California, and Corlett-Rivera and Hackman (2014), who received 1,343 responses from students and faculty in the humanities and social sciences at the University of Maryland. These surveys pose questions to members of a target population in order to gauge their awareness of, use of, and attitudes about e-books of different types (i.e. scholarly monographs, edited collections, and reference works) vis-à-vis other kinds of written materials, especially print books. The surveys also collect demographic data from respondents as to their college, department, and status. This allows for potentially useful comparisons between subgroups in the population. For example, when the Maryland survey asked users to indicate what format they prefer for scholarly monographs (print, e-book, no preference, it depends), results showed that 41% of all respondents preferred print, including 44% of faculty and 40% of graduate student respondents. The next question asked their format preference for edited collections: faculty preferred print to e-books, 36% to 25%, but graduate students chose e-books over print, 37% to 31% (Corlett-Rivera & Hackman, 2014, p. 268). Although most questions in surveys are tied to multiple-choice answers, there are usually a few open-ended questions that allow respondents to elaborate on “it depends” and provide details about their experience with--and within—particular texts. For instance, regarding his preferred format for scholarly monographs, one Maryland respondent wrote that it “depends on the urgency that I am reading with and what my end goal is, i.e. research, paper writing, personal betterment” (Corlett-Rivera & Hackman, 2014, p. 270).

Diary-based studies, supported by interviews, can provide an even closer look at reading behaviors because subjects (often students) write down --or are supposed to

write down--their activities, giving some details about not only what they read, but also the context and purpose of their reading (i.e. preparing for classes, preparing for exams, reviewing texts for research, gaining specific information, or learning new topics). With knowledge of the students' assignments and the tasks they perform, the investigators are able to identify different reading practices or techniques applied to different tasks and subjects. In a diary-based study of 39 University of Washington Computer Science and Engineering graduate students attempting to use Kindle DX e-readers to accomplish their academic reading, Thayer and colleagues (2011) analyzed the meta-level relationship between reading tasks and associated reading techniques. Students recorded their academic and leisure reading activities including specific tasks that proved difficult to perform on the Kindle DX, such as marking up texts, using references, using illustrations, and creating cognitive maps. Thayer and colleagues then associated each task with specific reading techniques, or "styles," defined by A. K. Pugh (1978, pp. 52-55):

- **Receptive reading** - reading sequentially from beginning to end with little variation in pace, to find out what an author has to say;
- **Responsive reading** - active engagement with arguments in the text, with frequent changes of pace, pauses, rereading;
- **Skimming** - a quick read to overview the structure or content of a text to locate potentially useful information;
- **Searching** - looking in a general way for answers to a question;
- **Scanning** - searching for a specific word or phrase.

Non-academic and leisure reading of novels and short articles indicated *receptive reading***;** text markup indicated *responsive reading*; and using references and using illustrations indicated *skimming*. *Skimming* also was associated with creating cognitive maps, the way readers notice and remember the physical location of information within a text and its spatial relationship to other locations in the text as a whole (Thayer et al., 2011, pp. 2921-2924). The study concluded that electronic documents on the Kindle DX were well suited to **receptive reading**, **searching**, and **scanning**, but did not support **responsive reading** and **skimming** well at all.

Before there were digital texts and computer logs, it was nearly impossible to study natural reading behavior over many pages of text. It was obtrusive and even "creepy" (Marshall, 2009, p. 96). It was also seldom done (McKay, 2011, p. 204). Now, with user session logs, researchers are able to collect reading pattern data unobtrusively from a large number of users as they interact naturally with e-books.

Description of this Study

The idea for this log analysis project was inspired by 1) the recent availability of detailed EBL session logs of Purdue Libraries' users; 2) a research article by Dana McKay (2011), who was probably the first to analyze reading patterns in EBL user logs; and 3) the authors' longstanding interest in comparative use of academic library collections in different subject areas.

EBL (Ebook Library) is a large aggregator that provides e-books to many academic libraries. In 2011, Purdue University Libraries chose EBL as the provider for the e-book patron-driven acquisitions (PDA) plan. Coordinated through the library's

primary book vendor, YBP, the plan started with an initial pool of a little over 11,000 titles. Although librarians have selected and purchased some of these e-books, most of the titles came into the catalog through a patron-driven acquisitions (PDA) plan the libraries set up through its primary book vendor YPB, so no fees were paid until the e-books were used by patrons. The collection grew steadily and, by the end of February 2014, it reached nearly 33,000 titles. Users have opened one-quarter of the titles at least briefly. To open a title, users link from the catalog record to the e-book and arrive at a summary page that features the book's cover, bibliographic information, and, often, an abstract—this web page is not recorded in the user log. From here they click "read online" and arrive at an introductory page in the EBL online reader that displays the e-book, starting with its cover, a scrollbar on the right, and, on the left, a hyperlinked navigation menu based on the table of contents. There are navigation keys and a jump-to-page feature above the e-book image, as well as a Search function. There is also a Download button that allows the reader to download a pdf or ePub version of the title into Adobe Digital Editions.

In her article "A Jump to the Left (and Then a Step to the Right): Reading Practices within Academic Ebooks," McKay (2011), a librarian at Swinburne Institute in Melbourne, Australia, pioneered the use of EBL logs to gain insight into users' e-book reading patterns, specifically those patterns associated with in-book navigation and with document triage or book selecting, that is, when a user chooses to elect or reject a book. She tracked sequential forward patterns and backward jumps, and verified that continuous sequential reading, the linear pattern associated with immersive reading of novels, seldom occurs for long in academic e-reading before readers jump forward and

back to other sections of the e-book (pp. 207-208). Although we did not adopt her quantitative methodology or units of measure, we were inspired by her description of three reading patterns comprising various degrees of linear forward movements and backward and forward jumps: *linear progression* for logs that proceed forward in a more or less orderly reading fashion; *contextual confirmation* for those instances where the reader makes a large jump forward in the paging and then backs up a few pages to verify the context of the part they then proceed to read continuously for several pages*;* and *exploratory assessment* for when the reader makes large jumps forward and backward in the pages consulted, apparently in search of particular material.

In her 2011 study and in a follow-up article (McKay et al., 2012), McKay's focus was on e-book selection behavior. This is especially relevant to EBL users at Swinburne Institute and other institutions where, after five minutes in an EBL "browse" session (with an un-owned title) or ten minutes (with an owned title), a window pops up that requires anyone who wants to continue reading to click "yes" on a dialog box and thereby initiate a "loan" session. Separating browse-session from loan-session data, McKay found statistically significant differences in the reading patterns in each group. Browse sessions showed more instances of *exploratory assessment* while loan sessions showed more patterns of *linear progression* and *contextual confirmation* (p. 19-20).

Separating browse and loan sessions was not relevant to the log analysis at Purdue because EBL users at Purdue do not have to take any action. The transition from browse to loan occurs seamlessly and users remain unaware of the change. Nevertheless, the authors were inspired by McKay's idea of analyzing EBL log data to

show reading patterns and hoped to devise a method to use EBL log data to support the hypothesis that users read and navigate within a book differently depending on the type (i.e., monograph or edited work) and subject area of the book.

This study reports on research analyzing data from EBL e-book user sessions at Purdue University to attempt to answer two questions:

1) How do users' reading practices differ when interacting with e-books that are authored monographs versus e-books that are edited collections of chapters by different authors?
2) How do users' reading practices differ when using e-books in different subject areas?

The authors expected the data to show significant differences, for example: that users would read authored monographs in a less jumpy and more continuous linear pattern than they read edited collections; that users of edited collections would proceed directly to one or two relevant chapters, rather than explore the whole book; or that users of animal science and technology e-books would do more searching and scanning than readers of history and literature e-books. The results of the log analysis, however, did not meet expectations. The similarities were more impressive than the differences, which were not as great as had been imagined.

Methodology

The analysis of logs of e-book use to describe reading behavior is still a new research venture. Consequently, the methodologies for this type of analysis are intuitive

rather than based on any theoretical considerations or on the results of past research. Some of the most important questions, such as, what are the basic “units” of analysis or how does one distinguish reliable data from dirty data, are still to be answered. Therefore, the methods used here are driven by the questions asked rather than by any previously established measurements or methodologies. From this perspective the results of this chapter are exploratory and descriptive, rather than conclusive as a comparative analysis. They are as much about the evolution of workable methodologies as they are about the results of the analysis.

For this study the authors pulled data from the EBL use report for Purdue University for one year (July 1, 2013 -- June 30, 2014). The resulting data set covered 29,884 user sessions with 5,245 titles viewed by 4,579 users. The user or reader session logs are part of EBL use reports available to Purdue through LibCentral, EBL’s administrative site, which collects detailed information on the use of EBL e-books. Although the data do not provide any personal details about each user, such as academic status or department, they do track each anonymized user’s e-book activities across time. Session details that were essential to this study include:

- duration of each session;
- page numbers in the sequence in which they were viewed;
- anonymized user identification for each user;
- EBL identification number for each title;
- bibliographic details for each title including ISBN and e-ISBN; and
- Library of Congress (LC) class and a broad subject heading for each title.

The report also provides names of author(s) and editor(s), but combines them without distinction within an author field, making it difficult to separate authors from editors. This difference is important in this study to distinguish the type of e-book used. To overcome this difficulty, the research team extracted the e-isbn from the EBL use report and then pulled matching title records from YBP's Gobi database that present author(s) and editor(s) in separate fields, and then merged these fields into the EBL use report.

EBL data come packaged as "user sessions" or "reader session logs." Each session log is a record of what transpires between the time the user opens the book and the time he or she stops reading. The same reader, however, may open and close the same e-book several times the same day or on immediately successive days. The authors decided that the best unit of analysis would be all of the reader session logs for the same reader while he or she was reading the same book. For simplicity we called this unit of analysis a **"Read."** This group of activities by the same person in the same book tells more about reading habits than does a single reader session log. Also, nearly all session logs show the reader flipping through pages numbered 1-5 when first opening an e- book. Although there are variations between e-books, these first few pages are invariably front matter, some of them being advertisements for other books or even blank pages. They contribute little or nothing to the analysis, so they were eliminated from the log. However, if a log began on a page number higher than five, say, page 15, then nothing would be eliminated.

It must be noted that the "page numbers" given in the log are ***file page image numbers*** rather than a book's ***real*** or ***actual page numbers*****.** For example, page 1 in

the log refers to the image of the book's cover, and page 15 in the log might refer to the image of the book's real page xii, a page in the introduction. It proved difficult for the authors to use a page image number from the log to find the equivalent actual page in an EBL e-book because the EBL online reader does not display image numbers. Patrons using the EBL online reader only see actual page numbers. An automatic way to translate or convert log image numbers into actual page numbers would make it easier to do research that combines log analysis with examination of e-book content. Fortunately, when EBL e-books are downloaded, the Adobe Digital Reader displays both real page numbers and file image numbers together. Because of the large number of reader logs, however, the authors did not include downloading e-books to the Adobe Digital Reader as part of the methodology.

Much of this log analysis focused on the sequence of page numbers for each Read. The objective was to find patterns that would indicate where the reader was going while looking at particular pages. Was the reader looking at consecutive pages, or was the reader jumping to later pages in the book or flipping back to earlier pages? Comparing sequences of page numbers between one Read in one title and another Read in another title is meaningless. So, in this study, to make comparisons possible the sequence of page numbers was converted to a sequence of page changes (i.e. Did the reader turn one page or did he or she jump ahead?). The sequences of page changes were then partitioned into units of **"reading passages"** where the partitioning was based on evidence that the reader had skipped over some reading material or had jumped back to earlier material. The word "jump" was used as part of the nomenclature to name these passages. (See the Appendix A for an illustration of the partitioning of the

page changes and naming them). The **"Passage"** itself contains a sequence of page changes that show that the user has read consecutive pages or skipped only one page or gone back only a single page. The rationale behind allowing one page skipped forward or one page turned back to be considered consecutive reading is that in a normal reading one sometimes comes across blank pages that are numbered or pages with illustrations that are numbered, or sometimes one turns back a page to see where one left off. The authors created five Passage distinctions:

1. **Forward (FOR):** A reading Passage that begins with no jumps.
2. **Small Jump Forward (SJF):** A reading Passage that begins with a forward jump of more than two pages but less than nine pages.
3. **Big Jump Forward (BJF):** A reading Passage that begins with a forward jump of nine or more pages**.**
4. **Small Jump Back (SJB):** A reading Passage that begins with paging back more than one page but fewer than nine pages
5. **Big Jump Back (BJB):** A reading Passage that begins with paging back nine or more pages.

The authors created small jumps and big jumps in order to distinguish between a pattern in which a reader examines pages that are near one another, probably within the same section of the book, and a pattern in which a reader examines pages that are far apart and probably in a different section or chapter. Nine pages, although somewhat arbitrary, seemed like a reasonable estimate of the average length of text that would fit within a section or chapter of a book.

For some analyses it was useful to join successive Passages into pairs of Passages. Figure 1 shows how a sequence of Passages, SJB BJF SJF BJB, is combined into pairs of Passages.

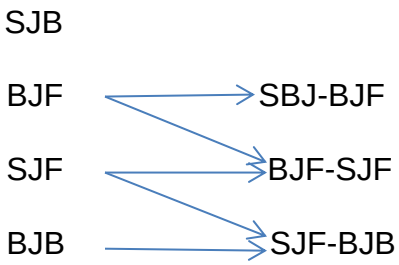

Figure 1. Transformation of a sequence of Passages into pairs.

These transition pairs provide another unit of analysis that allows us to see changes in the direction of turning pages; a simple count of Passage directions does not accomplish this. A sequence of Passages that jumped forward continuously and then backward continuously will give very different results than a sequence that is constantly alternating direction, even though the number of forward and backward jumps might be the same for both sequences. For a clearer understanding of these procedures we refer the reader to the Appendix A for an example.

Results

The EBL use report for Purdue University contained 29,884 reader session logs. The reader session logs pertained to 5,245 e-books read by 4,579 readers. There was a broad range of use of the e-books. For example, one title, *Handbook of Human Factors and Ergonomics*, was opened 1,551 times by 277 readers. In another example, a sinlge reader accounted for 1,664 reader session logs ranging over 703 e-books.

The 29,884 reader session logs reduced to 10,974 Reads. For in-depth analysis, the authors decided to select those Reads that had 11 or more pages and had one or more paired Passages. Those Reads with fewer pages or zero paired Passages did not provide sufficient data for in depth analysis. As a result, 7,224 Reads were analyzed in depth and 3,750 received only a summary analysis. Table 1 shows the basic data for both groups of Reads.

The data for the 3,750 Reads parallel closely the data for the 7,224 Reads. Looking at the data for all 10,974 Reads, we see that when readers jumped around in the text, it was more likely to be a jump backwards to earlier sections of the e-book. The number of times a reader turned pages back to an earlier section of the e-book (143,269) was greater than the number of times he or she jumped forward to a new section (97,571). This is confirmed by the number of Small Jump Back Passages (71,605) compared to the number of Small Jump Forward Passages (42,979) and by the number of Backward pairs (35,282) compared to the number of Forward pairs (14,797). Will these patterns repeat when subpopulations of the Reads are analyzed?

Table 1. Sums across Reads for different measures of e-book reading.

	Sums across all 10,974 Reads	Sums across 7,224 Reads of 11 or more pages	Sums across 3,750 Reads of less than 11 pages
Minutes	305,024	292,987	12,037
Pages	457,764	439,918	17,846
Sessions	29,884	24,439	5,445
Passages	179,780	172,469	7,311
Paired Passages	170,167	165,245	4,922
Individual Page Turns			
Consecutive turns	219,785	211,693	8,092
Jump forward turns	97,571	94,727	2,844
Jump back turns	143,269	138,909	4,360
Passages			
Forward Passages	7,903	5,914	1,989
Small jump forward Passages	42,979	42,178	801
Big jump forward Passages	26,526	26,017	509
Small jump back Passages	71,605	69,636	1,969
Big jump back Passages	21,812	21,500	312
Paired Passages			
Forward pairs	14,797	14,470	327
Alternating pairs	120,088	116,166	3,922
Backward pairs	35,282	34,609	673

The raw sums of data, however, do not reveal all. If one were to create distribution graphs for these data, they would be highly skewed, with a high number of Reads having low values and a small number with very high values. The distributions would replicate typical power law distributions. The cause for this type of distribution is that a large number of Reads were of short duration, only a few pages in length, while some Reads were extremely long. Any distribution of measures relating to how readers navigate the text will simply be a function of the length of the Read; means and

standard deviations will be uninterpretable. In order to compare Reads on the same scale, many of the measures for each Read were normalized by calculating ratios valued between 0 and 1 and then multiplying these ratios by 100. The result is a scale of 0 to 100 on which to compare data for individual Reads. In case the numerator and denominator are the same units, the result is a percent. The averages and standard deviations of normalized values will themselves be on the 0 to 100 scale. The length of the Read will have a small effect of the normalized values.

Limiting the analysis to Reads having a minimum of 11 pages and one pair of Passages gives sufficient data points for reliable insights into the reading patterns of academic e-books. The restricted set of 7,224 Reads included 3,424 e-books read by 3,580 different readers. Most Reads consist of one patron reading one e-book, but at the other extreme, the data reveal that one patron read 405 e-books and that one e-book was read by 260 patrons. The statistical data for this set of Reads are presented as averages, medians, and standard deviations for the general characteristics of the data and for the three units of analysis: page turns or jumps, Passages of page turns or jumps, and paired Passages. The reader is referred to Appendix B for a complete set of the statistics. In the discussion and analysis that follow the statistics in the tables are limited to those pertinent to the analysis.

Table 2 summarizes the principle features of patron reading habits for academic e-books. The average number of reader sessions for each Read was 3.4, and the average reading time spent on each Read was 40.6 minutes. The average number of pages read was 60.9 and the average number of Passages within those pages was 23.9. The medians for these three measures are lower than are the averages, showing

Table 2. Measures of reading patterns for 7,224 Reads.

	Averages	Medians	Standard Deviations
Duration of Reads in minutes	40.6	12.4	
Number of seconds to read a page	35.5		
Number of pages in a Read	60.9	37.0	
Number of sessions in a Read	3.4	2.0	
Number of consecutive pages turned	29.3	15.0	
Percent of consecutive pages turned	45.21%	43.50%	19.67%
Number of backward jumps in a Read	19.2	11.0	
Percent of backward jumps in a Read	32.77%	33.30%	13.25%
Number of Passages	23.9	14.0	
Number of paired Passages with forward jumps	2.0	1.0	
Percent of paired Passages with forward jumps	9.53%	7.10%	11.86%
Number of paired Passages with alternating jumps	16.1	9.0	
Percent of paired Passages with alternating jumps	72.00%	72.40%	16.19%
Ratio of # of Passages/# of pages in a Read	.409		.163

a skew toward the lower values in the series. On the other hand, the differences between average values and median values for the normalized variables are very small. The large number of Passages indicate a strong tendency to move about within the e-book. More striking was the high frequency of changing direction when going from one Passage to another: 72% of the paired Passages alternated between forward jumps and back jumps (e.g., BJF-SJB, SJB-SJF, etc.), while only 9.53% of such transitions maintained a steady forward reading direction, for example, BJF-SJF (see Table 2).

This suggests that academic e-book users are more engaged in skimming, searching, and responsive reading than in receptive reading. Nevertheless, on average 45.21% of the pages turned in a Read were consecutive pages. Keep in mind that users probably still spent more time actually reading these pages than performing quick jumps.

These broad statistical measures give a great deal of insight into the general patterns of patron reading behavior, but what might be the causes for such patterns? Do they come from different ways of constructing or formatting a text? Does the logical unfolding of concepts and explanations in different subject areas affect the way a book is read? Or are the causes basically determined by the different needs and objectives of the readers themselves? Given the data collected here, a random effects model would normally help answer such questions. It is doubtful, however, that the Reads are independent observations; in addition, the resulting model would have so many degrees of freedom as to minimalize its value. The less formal approach used here is to compare the averages of the normalized variables to understand any effects produced by e-book type and by different subjects, and to use the standard deviation of these variables as a surrogate measure for the effects of reader objectives.

One of the principle objectives of this study was to determine any differences in reading styles for authored monographic e-books and edited collection e-books. Table 3 shows that there are some small differences. Readers of edited collections tended to read more pages per book and to divide their progress through the book into more Passages. Dividing the number of Passages by the number of pages indicates that the

number of Passages is a function of the number of pages read. Overall there is great similarity in reading styles for both edited and authored e-books.

Table 3. Comparative measures of reading patterns for authored and edited e-books.

Type	Authored	Edited
Number of Reads	4,338	2,886
	AVERAGES	
Duration of Reads in minutes	40.4	40.8
Number of seconds to read a page	35.9	35.0
Number of pages in a Read	59.7	67.1
Number of sessions in a Read	3.2	3.6
Number of consecutive pages turned	29.0	29.7
Percent of consecutive pages turned	46.27%	43.61%
Number of backward jumps in a Read	17.5	21.8
Percent of backward jumps in a Read	31.45%	33.50%
Number of Passages	22.1	26.6
Number of paired Passages with forward jumps	1.8	2.3
Percent of paired Passages with forward jumps	9.54%	9.50%
Number of paired Passages with alternating jumps	14.9	17.9
Percent of paired Passages with alternating jumps	72.27%	71.59%
Ratio of # of Passages/# of pages in a Read	.404	.418

Another objective of this study was to determine whether or not there were significant differences between how books were read in different subject areas, or in different classes of the Library of Congress (LC) Classification. For this comparison the authors chose to analyze Reads in three large categories—humanities, social sciences, and STEM—and selected three groups of LC classes they thought would be not only representative of each category, but also would be different enough within each category that one would not replicate the other. As shown in Table 4, these LC classes,

which were drawn from the 7,224 Reads used for in depth analysis, formed a subset with a total of 3,907 Reads. Those Reads that fell into other LC classes were omitted.

Table 4. Number of Reads in each LC group for Reads used in analysis and Reads not used in analysis

Categories	LC (Subject)	Reads used in analysis	Reads not used in analysis	TOTAL	% reads not counted
Humanities					
	D & E (History)	277	136	413	33%
	PR & PS (English & American Literature)	184	120	304	39%
	N (Art)	69	29	98	30%
Social Sciences					
	L (Education)	424	198	622	32%
	HD, HE, HF, HG (Business)	633	353	986	36%
	PE (English Linguistics)	48	30	78	38%
STEM					
	QA (Mathematics)	153	61	214	29%
	SF (Animal Science)	1608	538	2146	25%
	T (Technology)	511	175	686	26%
TOTALS		3907	1640	5547	30%

The authors were concerned that the analysis for these subject areas might be skewed if some of the subject areas had a greater preponderance of very short Reads that would have been dropped from the analysis because they were part of the 3,750 Reads not analyzed in depth. To verify that this was not the case, the authors counted the number of Reads in each subject in both the analyzed and not analyzed groups. The 3,750 Reads not analyzed represent 34% of the 10,974 Reads. Table 4 shows the

percentages of Reads not analyzed for each subject area were all reasonably close to that 34%, indicating very little skewing of the analytical results.

There are fairly large differences between subject areas in the average times spent reading in an e-book and in the number of pages read (see Tables 5, 6, and 7).[1] Readers in all three of the STEM areas read on average more pages in an e-book than did readers in any of the humanities or social science areas. They also returned to the

Table 5. Comparative measures of reading patterns for three subject areas in the humanities.

	English & American Literature	**Art**	**History**
LC classes	PR & PS	N	D & E
Number of Reads	184	69	277
	AVERAGES		
Duration of Reads in minutes	34.3	32.0	49.1
Number of seconds to read a page	35.7	36.0	41.6
Number of pages in a Read	48.4	49.6	60.4
Number of sessions in a Read	3.1	2.5	3.3
Number of consecutive pages turned	21.1	25.6	29.5
Percent of consecutive pages turned	41.13%	48.37%	48.37%
Number of backward jumps in a Read	15.6	13.8	17.9
Percent of backward jumps in a Read	34.55%	30.47%	30.25%
Number of Passages	19.9	16.3	22.4
Number of paired Passages with forward jumps	1.5	1.1	1.9
Percent of paired Passages with forward jumps	8.40%	8.62%	9.35%
Number of paired Passages with alternating jumps	13.5	11.3	15.2
Percent of paired Passages with alternating jumps	72.63%	76.45%	72.51%
Ratio of # of Passages/# of pages in a Read	.441	.354	.395

same title for more reading sessions than did readers in any of the humanities or social science areas. On average, readers in the STEM areas also spent more time using an e-book than any group in the humanities and social sciences, except for historians.

Table 6. Comparative measures of reading patterns for three subject areas in the social sciences.

	Business	**Education**	**English Linguistics**
LC classes	HD HE HF HG	L	PE
Number of Reads	633	424	48
	AVERAGES		
Duration of Reads in minutes	35.4	32.4	33.2
Number of seconds to read a page	35.0	34.2	41.5
Number of pages in a Read	59.7	54.1	45.5
Number of sessions in a Read	3.0	2.9	2.9
Number of consecutive pages turned	30.0	26.8	20.5
Percent of consecutive pages turned	38.34%	46.18%	43.36%
Number of backward jumps in a Read	16.9	15.6	14.0
Percent of backward jumps in a Read	30.08%	31.73%	31.49%
Number of Passages	21.8	20.0	18.9
Number of paired Passages with forward jumps	2.0	1.7	1.8
Percent of paired Passages with forward jumps	11.07%	9.54%	10.40%
Number of paired Passages with alternating jumps	14.4	13.4	13.0
Percent of paired Passages with alternating jumps	70.82%	71.79%	74.83%
Ratio of # of Passages/# of pages in a Read	.397	.405	.473

Looking at the number of Passages into which the pages are divided, one sees more or less the same pattern, the STEM subject areas exceeding the others. The

same can be said for the number of paired Passages, both more with a forward direction and more with a back and forth direction. This could indicate that that readers in STEM were doing a lot of searching and scanning in pursuit of cross-references. One interesting difference, however, is between mathematics with 15 paired Passages with alternating jumps, and technology with 25.4 paired Passages with alternating jumps. One must consider, however, the effect of the number of pages read on the number of Passages and direction pairs. The last line in the Tables 5, 6, and 7 shows the ratio of

Table 7. Comparative measures of reading patterns for three STEM subject areas.

	Mathematics	**Animal Science**	**Technology**
LC classes	QA	SF	T
Number of Reads	153	1,008	511
		AVERAGES	
Duration of Reads in minutes	44.1	39.5	50.2
Number of seconds to read a page	37.3	31.9	32.2
Number of pages in a Read	68.4	69.5	91.7
Number of sessions in a Read	3.5	3.6	4.6
Number of consecutive pages turned	34.9	30.9	39.1
Percent of consecutive pages turned	48.94%	43.60%	43.08%
Number of backward jumps in a Read	19.5	22.5	31.1
Percent of backward jumps in a Read	29.74%	33.89%	33.82%
Number of Passages	22.9	27.2	37.6
Number of paired Passages with forward jumps	2.3	2.2	3.0
Percent of paired Passages with forward jumps	10.57%	8.17%	9.35%
Number of paired Passages with alternating jumps	15.0	18.6	25.4
Percent of paired Passages with alternating jumps	71.78%	73.14%	70.85%
Ratio of # of Passages/# of pages in a Read	.385	.412	.421

the number of Passages divided by the number of pages read. Here we see that the ratios for mathematics and for technology are almost the same. Converting the other reading pattern measures to ratios or percentages also had the effect of reducing the differences between most subject areas, but it also highlighted the fact that the percent of pages continuously read is higher for mathematics, history, and art than it is for other areas. Although there are subject area differences, what is more striking is the degree to which they are all very similar, implying that readers' reading strategies of e-books differ to only a small degree for different subject areas.

On the other hand, there appears to be a fairly large difference in reading patterns produced by the different objectives of the readers. Table 2 shows standard deviations for several variables used to measure the navigation of e-books. The standard deviations range in value from 11.86% to 19.67%. Given the very small effects for both book type and subject matter, and assuming there are no other factors producing a significant effect, the variance here is best explained by differences in reader objectives.

Conclusion

From a physical frame of reference, reading a book consists of eye movement and turning pages. Within a mental frame, the reading of a book is the recognition of words, the absorption of meaning from the words, and reflection on the meaning. From a causal perspective, the mental frame drives the physical frame. The research problem is to connect the two frames.

Across disciplines and between differently formatted texts such as edited and authored works, there exist small but plausible differences in a few of the basic measures for turning pages and spending time on the text. Perhaps just as striking is the degree of similarity between readers of e-books in different disciplines or subject areas. Yet the data show that individual users are different from each other in large ways in their reading patterns. Thus, the inference from the physical act of turning pages to the mental actions of the reader is that personal objectives are of greater importance for determining the physical reading patterns than is the nature of the subject material being read.

While the log data show that in general readers in general spend time engaged in continuous page-by-page reading--on average over 45% of pages turned were consecutive-- there was a surprisingly high percentage of transition pairs alternating between forward and backward jumps. This seems to indicate that academic e-book users are more engaged in responsive reading, skimming, and searching than in receptive reading. The differences between reading patterns in different subject areas conform to our intuitive understanding of how scholars absorb information and reflect on it. Historians, linguists, and mathematicians spend more time per page than do readers in the other disciplines. Traditionally, these are areas that require more concentration on textual details and reflection. The course requirements in different disciplines also certainly influence students' selection of particular texts and how they use those texts. Many of the most heavily used e-books, such as *Handbook of Human Factors and Ergonomics*, were in the STEM disciplines and were undoubtedly assigned readings or essential reference works for one or more courses. These practical concerns probably

explain why there were so many user sessions with these books and that the average time per page read was relatively short.

Although a powerful tool for revealing reader behavior patterns from many user sessions and large quantities of data, log analysis cannot provide insight into users' various circumstances and purposes. Future reading log analysis research should be informed by or done in tandem with the kind of survey or diary-based studies that gather information on readers' thoughts and intentions. Future e-book research should also be able to track or examine the specific content of e-book pages and connect the content to observed reading behaviors. Together these studies can lead to a more comprehensive understanding of reader behaviors.

Note

1. Statistical hypothesis testing was not used. Given that the reader session logs are not independent of each other, the meaning of such tests would be problematic.

Acknowledgments

The authors thank Rebecca Richardson, Electronic Resources Librarian, Purdue University Libraries, for assistance in gathering data from EBL and YBP.

References

Corlett-Rivera, K., & Hackman, T. (2014). E-book use and attitudes in the humanities, social sciences, and education. *portal: Libraries and the Academy 14*(2), 255–-286. http://dx.doi.org/10.1353/pla.2014.0008

Levine-Clark, M. (2006). Electronic book usage: A survey at the University of Denver. *portal: Libraries and the Academy,6*(3), 285-299. http://dx.doi.org/10.1353/pla.2006.0041

Li, C., Poe, F., Potter, M., Quigley, B., & Wilson, J. (2011). UC Libraries academic e-book usage survey. *CDL Staff Publications*: California Digital Library, UC Office of the President. Retrieved from http://www.escholarship.org/uc/item/4vr6n902

Liesaputra, V., & Witten, I. H. (2008). Seeking information in Realistic Books: A user study. *Proceedings of the 8th ACM/IEEE-CS Joint Conference on Digital Libraries* (pp. 29-38). New York: ACM Press. http://dx.doi.org/10.1145/1378889.1378896

Manguel, A. (1996). *A History of reading*. New York: Viking.

Marshall, C. C. (2009). *Reading and writing the electronic book*. San Rafael, CA: Morgan & Claypool.

McKay, D. (2011, November 28 - December 02, 2011). *A jump to the left (and then a step to the right): Reading practices within academic ebooks.* Paper presented at the OzCHI '11 Proceedings of the 23rd Australian Computer-Human Interaction Conference Canberra, Australia (pp. 202-210). http://dx.doi.org/10.1145/2071536.2071569

McKay, D., Hinze, A., Heese, R., Vanderschantz, N., Timpany, C., & Cunningham, S. (2012). An exploration of ebook selection behavior in academic library collections. In P. Zaphiris, G. Buchanan, E. Rasmussen & F. Loizides (Eds.), *Theory and practice of digital libraries* (pp. 13-24). Heidelberg: Springer. http://dx.doi.org/10.1007/978-3-642-33290-6_2

Nicholas, D., Rowlands, I., Clark, D., Huntington, P., Jamali, H.R., & Ollé, C. (2008). UK scholarly e-book usage: A landmark survey. *Aslib Proceedings: New Information Perspectives, 60*(4), 311-334. http://dx.doi.org/10.1108/00012530810887962

Pugh, A. K. (1977). Methods of studying silent reading behavior. *Research Intelligence, 3*(1), 42-43. http://dx.doi.org/10.1080/01411927700301115

Pugh, A. K. (1978). *Silent reading: An introduction to its study and teaching*. London: Heinemann.

Thayer, A., Lee, C. P., Hwang, L. H., Sales, H., Pausali, S., & Dalal, N. (2011). The imposition and superimposition of digital reading technology: The academic potential of e-readers. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp.2917-26). Vancouver, B. C.: Association for Computing Machinery, 2011). http://dx.doi.org/10.1145/1978942.1979375

Appendix A

Transforming Page Data for a Single Read

The sequences of page numbers for two reading log sessions of the same book by the same reader are:

Session one: 1,2,3,1,2,3,7,9,8,5,11,12,21,22,13,14,15

Session two: 16,17,18,19,20,23,24,44,45,33,34,35,31,32,49,50,48,47,51,52, 46,3,4,5,1,2,1,3

The sequences of page numbers with beginning page numbers 1 –5 removed:

Session one: 7,9,8,5,11,12,21,22,13,14,15

Session two: 6,17,18,19,20,23,24,44,45,33,34,35,31,32,49,50,48,47,51,52, 46,3,4,5,1,2,1,3

1. Concatenate page numbers into a single sequence or “Read”:

7,9,8,5,11,12,21,22,13,14,15,16,17,18,19,20,23,24,44,45,33,34,35,31,32,49,50,48,47,5

1,52,46,3,4,5,1,2,1,3

2. Convert page numbers to page changes:

2 -1 -3 6 1 9 1 -9 1 1 1 1 1 1 1 3 1 20 1 -12 1 1 -4 1 17 1 -2 -1 4 1 -6 -43 1 1 -4 1 -1 2

3. Partition page changes into “Passages”:

{2 -1} {-3} {6 1} {9 1} {-9 1 1 1 1 1 1 1} {3 1} {20 1} {-12 1 1} {-4 1} {17 1} {-2 -1} {4 1} {-6}

{-43 1 1} {-4 1 -1 2}

4. Add names to Passages. *Note.* The first number of each Passage is the number of pages jumped. The second is the number of continuous pages read after the jump.

FORWARD(2)(2) SMALL JUMP BACK(-3)(1) SMALL JUMP FORWARD(6)(2) BIG JUMP FORWARD(9)(2) BIG JUMP BACK(-9)(8) SMALL JUMP FORWARD(3)(2) BIG JUMP FORWARD(20)(2) BIG JUMP BACK(-12)(3) SMALL JUMP BACK(-4)(2) BIG JUMP FORWARD(17)(2) SMALL JUMP BACK(-2)(2) SMALL JUMP FORWARD(4)(2) SMALL JUMP BACK(-6)(1) BIG JUMP BACK(-43)(3) SMALL JUMP BACK(-4)(4)

5. Create a sequence of binary transitions or paired Passages:

FOR-SJB SJB-SJF SJF-BJF BJF-BJB BJB-SJF SJF-BJF BJF-BJB BJB-SJB SJB-BJF BJF-SJB SJB-SJF SJF-SJB SJB-BJB BJB-SJB

APPENDIX B

Averages, Medians, and standard deviations for raw values and normalized values for 7,224 Reads

	AVERAGES		MEDIANS		STANDARD DEVIATIONS	
N = 7,224 Reads	Raw Values	Normalized Values	Raw Values	Normalized Values	Raw Values	Normalized Values
Minutes	40.55		12.38		85.57	
Pages	60.89		37.00		80.65	
Sessions	3.38		2.00		3.94	
Passages	23.87		14.00		31.92	
Paired Passages	22.87		13.00		31.92	
INDIVIDUAL PAGE TURNS						
Consecutive turns	29.30	45.90%	15.00	43.50%	43.87	21.50%
Jump forward turns	13.11	22.90%	8.00	22.20%	17.90	11.40%
Jump back turns	19.22	32.90%	11.00	33.30%	27.26	16.30%
PASSAGES						
Forward Passages	0.81	8.10%	1.00	5.50%	0.38	8.70%
Small jump forward Passages	5.83	21.00%	3.00	22.20%	8.93	10.80%
Big jump forward Passages	3.60	13.60%	2.00	13.90%	5.42	9.50%
Small jump back Passages	9.63	37.00%	5.00	38.20%	13.80	11.30%
Big jump back Passages	2.97	10.40%	1.00	10.50%	4.95	8.50%
PAIRED PASSAGES						
Forward pairs	2.00	9.50%	1.00	7.10%	3.23	11.80%
Alternating pairs	16.08	72.00%	9.00	72.40%	22.56	16.20%
Backward pairs	4.79	18.40%	2.00	18.10%	7.53	14.30%