A Circulation Analysis of Print Books and E-Books in an Academic Research Library

Justin Littman and Lynn Silipigni Connaway

Over the past several years, a large number of libraries have begun to offer electronic books (e-books) to their patrons. The e-books provided to patrons are generally digital versions of books that also appear in print. They provide the same content as the print books, but are delivered in a different format. E-books offer a number of advantages over their print counterparts for both the patron and the library. For the patron, e-books offer 24/7 availability, remote access outside the physical library, full-text searching, and copying and pasting of text and images. For the library, e-books require no shelf space or reshelving, and are never lost, damaged, stolen, or overdue.

Despite these advantages, e-books still must prove their value to collection development librarians in one key respect—do patrons use them? Understanding whether patrons use e-books is important because collection development librarians generally take the usage of materials in the collection as an indication that the library is satisfying patron information needs. Thus, a complete assessment of the value of e-books in libraries requires examining the usage of e-books. However, evaluation of e-book usage is most likely to be useful when placed in the context of print book use, as print book usage provides something against which to measure e-book use. Comparing print book and e-book usage is appropriate when they provide the same content in both formats, particularly because librarians are increasingly faced with deciding whether to supplement or supplant new print book purchases with e-book purchases.

Assuming the data indicate that patrons do use e-books, then integrating them into a library’s collection development strategy requires understanding how they are used relative to their print counterparts. Relevant questions for collection development include: Are the same titles used in e-book format as in print? Does the overlap in usage vary by subject, or is it consistent across all subjects? Does the availability of e-books affect the use of print books? Addressing these questions about the use of e-books and print books will suggest implications for e-book collection development.
The purpose of this study was to perform a circulation analysis of e-book and print versions of the same titles at an academic research library. For example, the hardcover version of MIT Press’s *The Radiance of France* is compared against the e-book version of *The Radiance of France*. This study analyzed the use of netLibrary e-books and print books by the patrons of Duke University Libraries during an average study period of sixteen months per title. Duke University Libraries is an academic research library housing 4.5 million volumes and serving approximately 10,500 undergraduate and graduate students. Since 2001, Duke University Libraries has been offering e-books to patrons using the services of netLibrary, a division of OCLC Online Computer Library Center. At the time of this study, the netLibrary is an e-book service provider that serves the institutional market. At the time of this study, the netLibrary collection of e-books contained approximately 50,000 titles from more than 300 publishers. Individual institutions or consortia purchase e-books from netLibrary. netLibrary hosts the e-books on behalf of the institutions, making them available to those institutions’ patrons.

Previous Research

Circulation analysis is one of the traditional approaches taken to use studies and collection evaluation in libraries. The results of circulation analyses have been applied to a number of important issues, including evaluating collection acquisition policies, guiding such management decisions as allocating physical space for materials, identifying materials for offsite storage, allocating funding for materials, and suggesting approaches to deselection.

In the past, libraries collected circulation data by manually harvesting the circulation history from a book card or date label from a sample of books. The advent of library automation systems greatly simplified data collection. In these studies, a circulation represents an instance of a patron checking out a book for use outside the library. Circulations are generally for several weeks, but circulations can last several months in academic libraries. Two methods have been used to measure circulations: the length of time since last circulation and the frequency of circulation.

Circulation analysis assumes that the circulation of materials in a collection is an indicator of a library’s effectiveness. Or, as Wiemers Jr. et al. explain, high usage indicates that a collection is “good” since circulation is taken as evidence that a patron’s need is being met. In addition, practical applications of the results of circulation analyses assume that historical usage can be used to predict future usage.

One important weakness of circulation studies, as noted by Lancaster and Sumnerfield, Mandel, and Kantor, is that the methodology only reflects external circulations. That is, these studies do not account for in-library use of materials that does not result in a circulation. In comparing the circulation of print books and e-books, the California State University Electronic Access to Information Resources Committee and e-Book Coordinating Team accounted for in-library use of print books by determining the ratio of print book circulations to print books that were reshelved after in-library use. In the case of Henry Madden Library of Fresno State University, they found 1.03 in-library reshelvings for every circulation, prompting a doubling of print circulations to account for in-library use. However, since no widely accepted method for accounting for in-library use exists, in-library use is not corrected for in this study and represents a shortcoming in the methodology.

Circulation analysis generally is used to compare usage based on such variables as material age and subject area. Circulation analysis studies generally do not involve comparison of materials based on differences in formats—in particular, print versus electronic formats—as is being done in this study. An example of a study that does compare print and electronic formats is a 1998–2000 longitudinal study performed by Rogers at Ohio State University. The Rogers study used a survey to gather data on the frequency of use of print journals and electronic journals. The study only looked at journal usage in the aggregate, rather than the relative usage of individual journal titles (for example, the use of *College and Research Libraries* in print against its use in electronic format). Rogers concluded that electronic journal usage had increased, while print journal usage decreased. By the end of the study period, usage of both formats was roughly equal.

Several recent studies have compared usage of print and e-book versions of the same title. They use the traditional measure of circulations for print usage, and they use the measure of accesses for e-book usage. In general, an access is a single episode of a patron viewing an e-book.

One of the most comprehensive studies was the Columbia University Online Books Evaluation Project. As part of the study, 105 nonreference e-books and six reference e-books that were available in print format were made available to Columbia University patrons. (Not all of the titles were available in e-book format for the entire four-year study.) Data were collected on circulations of the print books and accesses of the e-books between winter 1995 and autumn 1999. Summerfield, Mandel, and Kantor concluded that for both the reference and nonreference titles, the e-books were used more than the print versions of the same titles. For example, “In spring 1999, nearly three times as many scholars clicked on the average online monograph book as circulated its print version.” While the Columbia University Online Books Evaluation Project covered a longer time period for some titles, the number of titles was much smaller than in the current study.
Four recent studies have been performed at the California State University Libraries, the University of Rochester, the University of Pittsburgh, and Wayne State University using data for netLibrary e-books. The netLibrary collection is a unique candidate for study because it is the largest collection of recent scholarly e-books available, with usage data from more libraries and covering a longer period of time than is obtainable from other e-book service providers.

The most comprehensive study using data for netLibrary e-books was completed by the California State University Libraries Electronic Access to Information Resources Committee and Coordinating Team. The Coordinating Team determined that 897 (60.1 percent) of the 1,492 e-book titles in the study were also available in print. In 2001, these print titles circulated 741 times. Between March 2001 and December 2001, e-books were accessed 1,039 times, which was annualized to 1,385 e-book accesses. By dividing the total number of accesses and circulations by the study set size, the e-Book coordinating team concluded that e-books had a 92.8 percent usage rate, whereas print books had a nearly identical usage rate of 92.4 percent, leading them to conclude that “the primary finding of this study is that when titles were available in both electronic and print formats, both formats were used.” The e-Book coordinating team noted:

... the Fresno campus has traditionally been print-oriented. Since the eBooks are not only new to the Fresno community but a new concept for most members of that community, there may be a lag time between introduction of the resources and their use. As more members of the community become familiar with eBooks, their rate of use is likely to rise.16

However, the e-Book coordinating team also noted that given the incongruity between e-book accesses and print book circulations, “the use of eBooks is most likely over-represented and the use of print books is under-represented,” and “in-house use raises the use level of the print books beyond that of the eBooks.”17

The e-Book coordinating team’s study has several weaknesses. First, the study set includes all e-books that were available to Fresno patrons, not just e-books that were available in both print and e-book. Since a committee representing the entire California State University library system selected the e-book collection, it is likely to contain titles that are not appropriate for Fresno patrons (and hence are unlikely to be accessed by Fresno patrons). Second, concluding that e-books had a 92.8 percent usage rate and print books had a 92.4 percent usage rate is misleading. This suggests that 92.8 percent of e-book titles were accessed, and 92.4 percent of print books circulated. It is entirely possible, however, that the accesses and circulations were accounted for by a small number of titles. The appropriate conclusion is that e-books were accessed .928 times per title and print books were circulated .924 times per title. This weakens the justification for concluding that there was heavy use of both e-books and print books.

In a much smaller 2001 study of e-book usage, Gibbons found that of the ten netLibrary e-book titles most frequently accessed by University of Rochester patrons, the University of Rochester libraries only owned one of those titles in print. This title circulated thirteen times over its lifetime in paper and was accessed 310 times in spring 2001 as an e-book.

Also in 2001, Connaway conducted a pilot study with the University of Pittsburgh using the same general methodology as the study repeated here. During a four-month study period, each netLibrary e-book title was accessed 3.7 times on average, while each print book circulated 1.4 times on average. Thirty percent of e-book titles were accessed at least once, while 10 percent of print titles circulated. While demonstrating that a study comparing the usage of print and e-book versions of the same title could address some interesting questions, the short time frame of this pilot study prevents drawing strong conclusions.

Though not a circulation study, Sutton’s recent report on Wayne State University Libraries’ experience using netLibrary’s Patron Driven Access (PDA) model for e-book acquisition has obvious implications for e-book collection development. In the PDA model, a library’s patrons have access to a large collection of e-books. However, the library only purchases e-book titles that have been accessed a certain number of times by the library’s patrons. (This differs from the standard purchase model in which a library purchases copies of e-book titles prior to any patron use.)

According to Sutton, during the study period e-book titles purchased via PDA averaged 4.12 accesses per title versus .43 accesses per title for traditionally selected e-books. In addition, collection development coordinators at Wayne State concluded that 92 percent of the e-book titles purchased via PDA were appropriate for the collection.

Method

The first step in this study was to match print books to e-books. MARC records for netLibrary e-books were extracted from the Duke University Libraries’ online catalog. The Duke University Libraries’ online catalog contained records for 14,398 e-books. Using Z39.50 queries against the catalog, e-book MARC records were matched with print book MARC records. Matches were made based on the International Standard Book Numbers (ISBNs) con-
tained in the 020$a subfield of the print book MARC records and the canceled ISBNs contained in the 020$z subfield of e-book MARC records. At the time of this study, cataloging practice for netLibrary e-books was to move the print book ISBN from 020$a to 020$z when creating an e-book MARC record from a print book MARC record. A small number of records were excluded when there was a single print book MARC record for a multivolume set, but there was a separate e-book record for each volume. In addition, records were excluded when the print record was added to the Duke University Libraries’ online catalog after the corresponding e-book record was added. After exclusions, 7,880 e-book and print book matches were used in the study.

After print books and e-books were matched, the second step was to obtain usage data for the print books and e-books. Circulation statistics were extracted from log files from the Duke University Libraries’ circulation system based on the local control number in the 001 field of the print book MARC records. Access statistics were extracted from netLibrary’s usage tracking system based on the netLibrary book identifier in the 035 field of the e-book MARC records. The measure of usage for netLibrary e-books is an “access.” In the netLibrary system, an access takes two forms. First is a “browse,” in which an e-book is in circulation as long as the patron is using the e-book. Once the patron stops using the e-book, it becomes available to another patron. Second is a “checkout,” in which an e-book is in circulation to a patron for a designated period of time. Compared to typical print book circulation periods, both browses and checkouts are for relatively short periods of time—from minutes to several days.

Once access statistics were extracted, the third step was to perform an analysis on the usage data. Previous studies involved a direct comparison of the frequency of print circulations and e-book accesses. While in some ways similar measures, print circulations and e-book accesses are also fairly incongruous. Circulations tend to cover long periods of time, whereas accesses cover short periods of time. In a single print circulation, a patron may use that book multiple times, whereas comparable use of a netLibrary e-book might involve multiple browses or checkouts. Also, e-book accesses include use both inside and outside the library, whereas circulations include only external use. Given these considerations and the relatively short time frame of the study period, this study adopted a different method from earlier studies by comparing whether e-books or print books had circulated or been accessed during the study period rather than the frequency of circulations and accesses. The study period was defined as the period between the e-book MARC record being added to the Duke University Libraries’ online catalog and August 2002. Thus, each title had a different study period. The first records were added in February 2001, though the average study period was sixteen months. This method partly, though not entirely, accounted for the incongruity between print circulations and e-book accesses.

To analyze usage by subject area, each title was assigned to one of thirty subject categories based on its Dewey Decimal classification. Subject areas include literature, philosophy, psychology, computers, arts, technology, engineering, and manufacturing. This approach was used to identify accesses of both the print books and the e-books by subject areas.

To analyze the effect of the availability of e-books on print book circulation, the circulation of print titles was compared for the year prior to the availability of the e-book against the year following the availability of the e-book. This was accomplished by identifying the set of print books that were available for at least one year prior to the addition of the corresponding e-book’s MARC record to the Duke University Libraries’ online catalog. This data set included 7,456 print books. Comparisons were then performed between the circulations in the year prior to the addition of the e-book MARC record and the circulations in the year after the addition of the e-book MARC record.

**Results**

Of the 7,880 titles that were available in print and e-book, 3,158 e-book titles were accessed and 2,799 print titles were circulated during the study period. In print and e-book format, 1,688 titles were used. In e-book format, but not in print, 1,484 titles were used. In print, but not e-book format, 1,125 titles were used. In either format, 3,597 titles were unused. The results for the titles that were used in either format are represented in figure 1.

![Figure 1. Titles used in print or e-book format](image-url)

Though not appropriate for direct comparison for the reasons given earlier, total and average circulations and accesses were computed. During the study period, print...
books circulated 6,998 times, averaging 0.88 circulations per print title or 2.48 circulations per print title that circulated. E-books were accessed 10,821 times, averaging 1.37 accesses per title or 3.43 accesses per e-book title that circulated.

Results by subject were computed in two different ways. First, for each of the top subjects represented in the study set, the percentage of the titles in the overall study set by subject, the percentage of the titles used in e-book by subject, and the percentage of the titles used in print by subject were determined. These results are given in table 1.

Second, for each of the top subjects represented in the study set, the percentage of the titles in that subject used in e-book and the percentage of the titles in that subject used in print were determined. These results are given in table 2.

Results also were computed to permit the evaluation of the impact of the availability of e-books on the circulation of print books. There were 6,139 circulations of print books in the year prior to the introduction of the e-book. This decreased to 4,738 circulations of print books in the year after the introduction of the e-book. This represents a decline of 22 percent in print circulations. By contrast, total circulations at Duke University Libraries increased by 5.2 percent between the 2000–2001 academic year and the 2002–2003 academic year.

Of the 7,490 print titles available in the year prior to the introduction of the e-book, 1,571 titles circulated in the year before and the year after the introduction of the e-book. In the year before, but not the year after the introduction of the e-book, 1,149 titles circulated. In the year after, but not the year before the introduction of the e-book, 820 titles circulated, and 3,932 titles did not circulate in the year before or the year after the introduction of the e-book. The results for titles that did circulate are represented in figure 2.

**Discussion**

Based on this method of evaluation, e-books received 11 percent more use than comparable print books. Given their recent introduction to patrons at Duke, this suggests rapid growth in the adoption of e-books. The high usage of e-books relative to comparable print books at Duke supports the general findings of the Columbia University Online Books Evaluation Project study, the Gibbons study, and the Connaway study, which found heavier usage for e-books. In addition, the popularity of e-books may increase as patrons become more familiar with e-books and as Duke Libraries expand the collection of e-books available to patrons. Note, however, that any conclusions reached from this study should be tempered by taking into account the incongruity

<table>
<thead>
<tr>
<th>Subject</th>
<th>% of titles in study set</th>
<th>% of titles used in e-book</th>
<th>% of titles used in print</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social Sciences: General</td>
<td>16</td>
<td>17</td>
<td>18</td>
</tr>
<tr>
<td>Business, Economics and Management</td>
<td>16</td>
<td>15</td>
<td>14</td>
</tr>
<tr>
<td>Literature</td>
<td>13</td>
<td>11</td>
<td>13</td>
</tr>
<tr>
<td>History: World and General</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Philosophy</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Religion</td>
<td>5</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>History: United States</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Political Science</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Arts</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Education</td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Medicine</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Law</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Psychology</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Computers</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Other subjects</td>
<td>12</td>
<td>11</td>
<td>9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subject</th>
<th>% of titles used in e-book</th>
<th>% of titles used in print</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social Sciences: General</td>
<td>42</td>
<td>40</td>
</tr>
<tr>
<td>Business, Economics and Management</td>
<td>36</td>
<td>31</td>
</tr>
<tr>
<td>Literature</td>
<td>36</td>
<td>36</td>
</tr>
<tr>
<td>History: World and General</td>
<td>41</td>
<td>38</td>
</tr>
<tr>
<td>Philosophy</td>
<td>39</td>
<td>34</td>
</tr>
<tr>
<td>Religion</td>
<td>46</td>
<td>43</td>
</tr>
<tr>
<td>History: United States</td>
<td>27</td>
<td>28</td>
</tr>
<tr>
<td>Political Science</td>
<td>43</td>
<td>43</td>
</tr>
<tr>
<td>Arts</td>
<td>46</td>
<td>42</td>
</tr>
<tr>
<td>Education</td>
<td>38</td>
<td>31</td>
</tr>
<tr>
<td>Medicine</td>
<td>51</td>
<td>42</td>
</tr>
<tr>
<td>Law</td>
<td>34</td>
<td>34</td>
</tr>
<tr>
<td>Psychology</td>
<td>57</td>
<td>49</td>
</tr>
<tr>
<td>Computers</td>
<td>66</td>
<td>53</td>
</tr>
</tbody>
</table>

**Figure 2.** Circulating print titles, before and after introduction of e-book titles
between print circulations and e-book accesses previously discussed.

The titles that circulated in print overlap somewhat with the titles accessed in e-book format. Of the titles that were used in print or e-book, 39 percent were used in both formats. Thirty-four percent were used in e-book only, and 27 percent were used in print only, suggesting that some patrons may be using e-books and print books for different purposes; for example, an e-book for quick reference but a print book for intensive reading.

A fair amount of overlap also occurred between the titles that did not circulate in print and were not accessed in e-book format. Seventy-one percent of titles that did not circulate in print were not accessed in e-book format. This suggests that the same titles that were unpopular in print were also unpopular in e-book format.

While initially the high rate of titles not used in print (64 percent) and e-book format (60 percent) might seem alarming, one should remember that this study covers a short time frame. If this study was conducted over a time frame of at least several years, these rates can be expected to drop substantially.

When analyzed by subject, both e-books and print books were used relatively in proportion to their subject representation in the entire study set. So, for example, business, economics, and management represented 16 percent of the entire study set, 15 percent of the titles used in e-book, and 14 percent of the titles used in print. This suggests that from a subject standpoint, the study set was well suited for the information needs of Duke’s patrons since no subject received significantly more or less use than its representation in the collection. It also suggests that patron e-book usage may be dictated by the availability in electronic format of titles and subject areas.

This study indicates that most of the top subjects (social sciences, business, and literature) were used approximately as much in print as in e-book format. Titles in education, medicine, psychology, and computers were used more in e-book format than in print. An e-book collection development strategy that focused on these subjects of higher usage may provide maximum benefit.

Anecdotal evidence suggests that the availability of e-books increases the usage of print books, since some patrons use e-books for “sampling” prior to acquiring the print book. Similarly, the National Academy Press claims that making its e-books freely available has driven print sales. This study suggests that the availability of e-books leads to a decrease in the usage of print books. One likely explanation for this trend may be that e-books are satisfying the information needs of patrons, in some cases obviating the need to utilize the print book. Other possible explanations for this trend include the continued aging of the print collection or a general decrease in print circulations. One implication of this trend is that collection development librarians consider e-books for additional copies when the existing print copies receive heavy usage.

Conclusion

If, as was suggested earlier, proving the value of e-books requires demonstrating that patrons use e-books, then the preliminary evidence provided in this study suggests that e-books do provide value. Despite the recent introduction of e-books at Duke University Libraries, the use of e-books is already substantial relative to their print counterparts. Although this could be attributed to the incongruity between print circulations and e-book accesses, expanding the size and prominence of e-book collections in academic research libraries seems justified, combined with continuing study and usage comparison of the two formats.

In addition to demonstrating the usage of e-books, this study suggests some approaches to e-book collection development. Attention should be paid to titles that particularly benefit from additional functionality offered by an electronic format, such as reference books. In certain subject areas, such as the social sciences, e-books may provide more benefit (assuming usage is an indicator of benefit) than other subject areas. Lastly, e-books are excellent candidates for additional copy purchases when print copies of titles are receiving heavy use.

Generalizing the conclusions of this study requires performing e-book circulation studies in other comparable academic research libraries, in other types of libraries (such as small academic libraries and public libraries), for longer time frames, and with different types of e-book content (such as trade content). As librarians continue to acquire e-books, circulation analyses will become increasingly important to identify a more complete understanding of e-book usage patterns. Collection development librarians also can use these data to create e-book collection strategies and policies that better meet user needs.

References


16. Ibid.

17. Ibid.


22. To extend this methodology over a longer time frame, one might want to consider determining whether e-books had been accessed or print books had been circulated in a series of one-year periods.

