Using Logistic Regression to Examine Multiple Factors Related to E-book Use

Karen Kohn

Many studies have tried to identify factors that make electronic books (e-books) in academic libraries more likely to be used. For instance, are demand-driven acquisitions used more than titles in packages? Are e-books in the sciences used more than e-books on art? Most of these studies are limited to one or two variables. This study introduces logistic regression, which can incorporate multiple variables to determine which factors are the most useful in predicting e-book usage. The variables considered in this study are LC class, university press or other publisher, and platform. In the collection studied, the classes with the highest odds of being used were A (General Works), followed by F (History of the Americas), H (Social Sciences), and Q (Math and Science).

Karen Kohn (karen.kohn@temple.edu) is the Collections Analysis Librarian at Temple University in Philadelphia.

Academic libraries are struggling to understand the role of electronic books (e-books) in their collections. Not all potential book purchases are available electronically, and patrons frequently claim they prefer print. Yet, for reasons including appealing purchasing models, the desire to reach remote patrons, and evidence that e-books are used, libraries are increasingly buying e-books. The addition of this format to academic library collections raises the question of how to evaluate their usage. This is much more complicated than the parallel task of evaluating print book usage. Not only do subject matter, publication date, and publication type (e.g., reference book, conference proceedings, monograph, edited volume, etc.) affect usage (as for print), but e-books have a variety of user interfaces and are selected through a wider variety of methods. Like print, e-books can be selected through an approval plan or by firm order (i.e., a librarian selecting a specific book). They are often available to purchase as demand-driven acquisitions (DDA), evidence-based acquisitions, subscription packages, or as publisher collections. Open access e-books are also becoming available on several platforms, and libraries are adding these to their catalogs. Due to the range of selection methods, interfaces, and other characteristics, the variety of factors that affect whether an e-book gets used is much broader than those affecting print book use.

As libraries generally want to purchase items they expect will be used, many studies have attempted to identify factors that make e-books more likely to be used. Studies of e-book usage most often consider just one or two variables. For instance, are DDA more likely to be used than titles in packages? Do e-books in the sciences get used more than those for art? These questions are helpful, but the findings of such studies are only a beginning. A publisher package might

receive more usage than an aggregator package, making it seem as if the quality of the publisher drives usage, when in fact the publisher package might simply have more current books or more relevant material than the aggregator package. E-book packages differ in so many ways that it can be difficult to know which feature drives use. In a study comparing usage of netLibrary and Ebrary collections, Tucker notes that differences could be related to the age of books in each collection or to user preferences for a particular interface.[1] Slater similarly notes that, in his comparison of Safari and netLibrary, "It is not possible to definitively determine . . . if it is the contents of the collection or the presentation of the collection that motivated users to choose one . . . over the other."[2] Since each book has a variety of features that could influence usage, there is a need for research that can simultaneously consider multiple factors.

A useful way to see which variables are most strongly correlated with usage is to combine multiple variables in a regression equation. By putting several variables into an equation that predicts an outcome, regression allows the researcher to separate the effects of each variable. To contribute to the methodology of measuring e-book use, this paper presents a logistic regression model that correlates several variables with the predicted usage of e-books in a large academic library. The research question is this: is it possible to identify characteristics of an e-book that will predict whether it will be used? The variables considered here are Library of Congress (LC) Classification (as a stand-in for subject), platform, publisher type, and usage of comparable print books. Though some variables of interest could not be included in the study, most significantly selection method, the methodology used can provide a model for others to expand upon and contribute to existing literature that has reported on how usage varies according to subject and publisher type.

Literature Review

Factors Considered in Previous Studies

Probably the most common question asked in the literature on e-books is which disciplines receive the heaviest usage.[3] The questions asked range very broadly, however, leading Wilkin and Underwood to claim, "There is no well-defined and stable problem statement regarding the study of e-book usage."[4] The only nearly universal feature of research on e-books is that it almost always tries to correlate a particular feature of e-books with rates of usage. A variation of the question of which subjects receive greater e-book use is one that also considers print usage, asking which subjects show a greater preference for e-books over print or the reverse.[5]

Another issue is selection method. There is a wider variety of selection methods typically used for e-books than for print. For instance, e-books can often be purchased as packages from aggregators or publishers, librarians can select individual titles for one-time purchase, titles can be added to a library collection based on an approval plan or made available for patrons to select using DDA. E-books are also sometimes freely available as open access. Consortia may purchase e-book packages, and the individual library does not get to choose which titles are included. Carrico et al. studied whether the selection method of e-books predicts the level of use (i.e., do firm orders, DDA, or purchased packages get used the most often?)[6] Levine-Clark hypothesized that selection method was the source of the differences he noted between usage of books in EBL and Ebrary. In worldwide data, a higher percentage of EBL books were used than Ebrary books, which Levine-Clark attributed to the fact that libraries select their EBL holdings title-by-title, whereas Ebrary tends to sell its books as part of a subscription package.[7]

Some researchers have speculated that it is not the subject or selection method that explains which e-books receive the most use but rather the kind of publication. A common finding is that reference materials are more popular in e-book form than monographs.[8] Bucknell further subdivided the books in his study into the following types: monograph, proceedings, contributed volume, professional book, textbook, and reference, while Sullivan and Leach compared monographs to edited volumes.[9] Several authors compared whether university press books are used more than other types or have asked which publishers' books are more likely to be used.[10]

As mentioned earlier, some studies have noted difficulties in understanding which variable accounts for differences in usage. When comparing Ebrary and netLibrary, Tucker explained that the collections differ in both selection method and currency. The former is a subscription package whose contents can change periodically, while the latter is an older collection of firm orders, with select newer titles added.[11] A few studies have tried to tease out how different variables interact with each other.[12] Thus far, the field has concerned itself with more fundamental methodology questions such as how to classify e-books by subject and how to compare e-book and print use.

Methods Used in Previous Studies

Wilkin and Underwood lament the lack of a "research paradigm" for e-books. They state that "there is no consensus on how to reliably measure ebook usage," a complaint with which Fry concurs.[13] The field lacks standardized ways to compare print and e-book usage or standard ways to interpret electronic usage. Although COUNTER provides an

international standard of what elements should be included in a usage report, the standards still allow for widely differing ways of measuring the extent of use. Proprietary vendor reports can provide additional information of interest, such as the amount of time a user spends on a book, but this information is not available for all platforms.[14] Wilkin and Underwood also note that several studies exist that rely only on surveys, which reveal user preference rather than user behavior. Additionally, they point out that surveys related to e-books are particularly problematic given that users may not correctly understand some of the terminology used in the survey.[15]

Several authors note problems with using COUNTER reports to compare e-book usage between vendors. The COUNTER Code of Practice, Release 4, offers a report called Book Report 2 (BR2) that lists how many "sections" of a book were viewed within the reporting period. The instructions describe a section as "chapter, encyclopedia entry, etc.," and specify that the report should indicate what counts as a section.[16] In Release 5, similar data will be found in Book Report 1 (BR1), which will contain a field titled "Total Item Requests." The documentation for Release 5 explains that this number "will vary significantly based on how the content is delivered, indicating that item requests in Release 5 will be as difficult to compare across platforms as section requests are in Release 4."[17] What each vendor counts as a section varies widely. Bystrom offers a chart of thirteen e-book packages and how a section view is defined for each. The most common definition is a chapter, but several count each page viewed as a section view, and one counts every five pages.[18] Cox notes an e-book provider that counts each three pages as a section view.[19] For a reference work, a "section" could be simply a dictionary definition.[20] Even when section views are consistent between platforms, limits on simultaneous users can also lead to significant differences in usage, as a platform that limits simultaneous users will have fewer total section views than one that allows unlimited simultaneous users.[21] Moreover, the interface will affect whether certain actions are counted in the usage statistics. As Levine-Clark, Paulson, and Moeller point out, if a book's landing page includes a table of contents and a blurb, users might view that page and decide against viewing the book. If there is no landing page, users will access the book to see the table of contents, and usage reports will indicate that this book was used even if the patron decided not to read any further.[22] In addition, some interfaces provide easier downloading than others. A patron who downloads a book can return to the downloaded copy repeatedly without it counting as an additional use, whereas an interface on which downloading was difficult could encourage patrons to return to the online option and their usage will subsequently be logged each time.[23]

Due to inconsistencies between COUNTER reports, several people simply count whether a book has been used rather than the number of uses. Littman and Connaway were the first to classify books simply as used or not used, and this strategy has since been used by others.[24] Knowlton makes the case for this method by pointing out that so few books in his library's collections were used that the difference between used and not used books is significant, whereas differences in the amount of use each book receives is marginal.[25]

Counting whether a book has been used rather than how often it has been used not only alleviates the problem of inconsistency in COUNTER reports, but also facilitates comparisons between e-books and print. Knowlton observed that comparing the two formats is "nearly impossible" to do accurately.[26] Kimball, Ives, and Jackson assert that the "traditional comparison" is between print checkouts and e-book accesses, although they acknowledge that both of these measures are inaccurate.[27] It is well known that print circulation, the standard measure of usage, is not only a limited measure in itself but also measures something very different from what e-book usage represents.[28] Checkouts do not tell us how extensively users have read a book. They could have read it cover-to-cover or simply looked at a few pages. Loan periods also affect circulation counts, as a book that is borrowed by a faculty member who is allowed to check out books for a year will show less use than a similar book that was borrowed by an undergraduate for a month.[29] Additionally, circulation does not contain information on books that were used in-house or that someone glanced at and decided not to use. The latter use case is counted in e-book usage statistics. Not all of these problems are corrected by counting whether a title was used rather than the number of uses, but this is beginning to be recognized as the preferred method for comparing e-book use to print use.

When comparing e-book and print usage, one not only needs a comparable measure of use but similar sets of books. Several studies have used paired lists where each title is held by the library both in print and as an e-book.[30] Goodwin uses Duke University Press books as the basis of her comparison. Since the Press offers an option whereby a library that purchases the e-book collection can pay a small fee to also receive the print, some libraries own recently published books by Duke in both formats.[31] When there is not a known collection that is duplicated in both formats, another option is searching the library's e-book holdings against the catalog to find matching print book records.[32] This can be laborious, however, and result in a very small set of books, as often libraries will have a policy that says they do not routinely purchase the same titles in different formats.

Recent studies have developed strategies for finding similar *groups* of books to compare even when the titles are

not the same. Fry examined all the books acquired within the same time period, regardless of publication date.[33] Knowlton considered all print books acquired during a certain time period and compared these to e-books from the library's largest e-book collections. He also excluded print books that do not circulate.[34]

When sets of e-books and print books are selected for comparison, and information is collected on whether they have been used, there are still several ways to make the comparison. It is important not to simply look at the number of uses in a particular format without taking into account the size of the collection. Fry points out that if print use is declining, it may be because the library is buying fewer new books, whether due to decreased circulation or more economical purchasing options for e-books.[35] A fairly common measure that considers the collection size is Percent of Expected Use, or PEU. Mills coined this term in 1982, and it has subsequently been used in several studies.[36] PEU represents the percentage of all usage from a particular subset of the collection divided by the percentage of the full collection making up that subset. For example, if history books are 20 percent of a library's holdings, but only 15 percent of that library's total circulation is from history books, the PEU for history would be 15 percent ÷ 20 percent, or .75. PEU can be measured for either print books or e-books, and since the units are the same regardless of format, comparisons can be made between the PEU for the same subject in both the print collection and the electronic collection. Knowlton calculated the difference between each subject's PEU for print and for electronic, as an indicator of the degree of preference for one format over another.[37] Slater asked whether the two PEUs are correlated. He found a positive correlation between print PEU and electronic PEU by subject, meaning that subjects with heavy usage in print also receive heavy usage in e-books.[38]

Dividing e-books into subject categories raises another methodological question, which is how to obtain subject classification information for e-books. COUNTER reports do not include call numbers, and MARC records provided by an electronic resource management system (ERM) do not always include call numbers for e-books. Some studies used vendor-provided subject categories, which do not correspond with LC classes or subject headings.[39] This makes it difficult to compare usage from one collection to other collections. Tucker compared books from netLibrary and Ebrary, which at the time of his study offered LC call numbers in their reports.[40] Carrico et al. mention using proprietary vendor reports for the benefit of call numbers provided therein.[41] Studies that use paired lists of titles, where each book is owned both electronically and in print, can use the print record's call number.[42] If the catalog records include call numbers, it is possible to match the ISBNs from a vendor's usage report to catalog data to pull in the call

numbers.[43] In studies that match call numbers with books, the call numbers are commonly mapped to the institution's programs, and the program becomes the unit of analysis.[44] Another option is to use the LEFT function in Microsoft Excel to create a column that lists only the first letter of each LC classification number, which can then be treated as a category that roughly corresponds with a discipline.[45]

Findings of Previous Studies

As stated earlier, the most common question about e-books is which subjects are most used. This is sometimes a simple question of comparing subjects to each other within one set of usage data and other times is framed as which subjects have the strongest preference for e-books over print. Answering the former question, Slater found that the most-used subjects in his library's netLibrary package were math and science.[46] Knowlton's study found that e-books in the general social sciences, psychology, and education had the highest PEU.[47] In Sprague and Hunter's collection, titles related to agriculture, botany, geology, and biology were the most likely to be used, with a surprisingly high rate for art. Anthropology and chemistry also had high rates of usage.[48]

With studies that compare e-book and print use, sometimes subjects with high e-book use have been heavily used in both formats, while other subjects are strongly preferred in one format over the other. Knowlton found social sciences to be a popular subject in e-books, though even more popular in print.[49] Slater, in contrast, found that math and science were the most popular subjects in netLibrary. Usage analysis revealed that these subjects also showed a preference for online over print. Users seeking books on technology, engineering, media, and communications also preferred e-books, while the subjects with the strongest preference for print were world history and language and linguistics.[50] Littman and Connaway also found that users preferred education, psychology, computer science, and medicine e-books over print.[51] Christianson and Aucoin found the strongest preference for print was with history books.[52] As these findings vary between institutions, additional research might clarify whether there are common trends regarding which subjects are used more in e-book form or if each institution needs to measure locally.

In addition to comparing e-book and print use, some studies have asked how the two relate to each other. Slater tested a correlation between print book use and e-book use by subject and found a moderate correlation between the two, with subjects that were heavily used in one format also being heavily used in the other.[53] Christianson and Aucoin found a positive but very low correlation at the individual book level, i.e., a print book that was used was slightly more likely to be used in the electronic form.[54] Littman and Connaway reached a similar conclusion: books used in print

frequently were used electronically.[55] Sullivan and Leach asked whether e-books might serve a discovery function, letting users skim a book that they would later decide to borrow in print for more in-depth reading.[56] They concluded that this was not the case, though Hobbs and Klare's small-scale qualitative research project found that students use e-books to determine what they want to read and then obtain the print for lengthier reading.[57] Littman and Connaway similarly suggest that e-books do not promote usage of their print counterparts, and in fact, in their study, print books were less likely to circulate once an e-book edition became available.[58] Others try to pinpoint whether the different formats serve different needs. For instance, are electronic materials more popular at a particular time in the semester, such as during finals when a student might be working close to a deadline and not have time to go to the library?[59]

Other studies have considered whether university press books receive more use than other books and if specific features of interfaces correlate with usage. Christianson and Aucoin found university press books to be more popular in print than as e-books, but these were still less likely to be used in either format than other books.[60] They speculate that this may be due to the specialized nature of university press publications and to the fact that they are usually meant to be read in a linear fashion that is more suited to print. Levine-Clark and Paulson found the opposite—that university press e-books were used more than other e-books.[61] They attribute this to the fact that university press books are of higher quality than trade publications. Surveys have reported various stated preferences for certain characteristics of e-books such as the ability to print, download for offline reading, or copy and paste text.[62] To this author's knowledge, no studies examined whether users' behaviors correspond with these stated preferences.

Method

This research was conducted at Temple University, a large institution with a Carnegie Classification of Highest Research Activity. The university libraries provide access to more than a million e-books, including an aggregator collection, several publisher packages, open access collections, and subject-specific packages. The main library has had a DDA program since July 2014.

The present study considered factors similar to those that have been studied previously, and introduces a methodology that enables several variables to be simultaneously considered. Like the studies described earlier, this study considers the subjects of books to see which receive the most use and why. It also takes into account whether a book is published by a university press and platform differences, and seeks a relationship between print usage and e-book usage for each subject. Some other variables that would have been desirable to consider are type of book (reference, monograph, edited volume, textbook, or other), selection method (DDA, firm order, or package), and various interface features, such as whether there is a table of contents landing page and if books are indexed in Google. It was not possible to include these variables because the largest collections of e-books in the author's library are not reference, nor do they have significantly different selection methods. Indexing in Google was hard to measure in a standardized way. It is hoped that the methodology used here can be expanded in future studies to include additional variables.

The e-book collections used in this study are Ebrary (Academic Complete collection), MyiLibrary (a mix of DDA and firm order titles), netBASE (engineering collection), Springer (publisher complete collection), and Wiley (publisher evidence-based acquisitions collection). After the research was completed, the library's holdings in both Ebrary and MyiLibrary were migrated onto the EBook Central Platform. The analysis and discussion here refer to the platform that hosted the e-books during the time period for which usage was being measured. The sample consisted of all titles published in 2015 from each of the above-mentioned collections. There were two reasons for using samples rather than the full holdings. One is that the smaller subset was a more manageable number for looking up call numbers. Using only books from 2015 simplified the analysis by avoiding the question of whether to consider the age of the book and the acquisition date when looking at usage. The platforms studied were the five largest platforms for which the library had access to books published in 2015.

For each of these collections, the title list was downloaded and ISBNs were pasted into OASIS, ProQuest's online ordering and tracking tool, to obtain call numbers. OASIS allows users to paste long lists of ISBNs into a search box, and in this case, five hundred to a thousand were pasted at once. The resulting list was exported to Excel, and the call numbers from the export were copied and pasted into the title list. A small number of titles lacked call numbers in OASIS and were removed from the sample. In the Excel document containing the title lists, a new column was created containing only the first letter of each call number, so that each book was assigned a single-letter LC class.

The variable for publisher type (university press or other) was assigned by filtering the title list for rows with the word "university" in the publisher field. It would have been desirable to create more categories of publishers, such as scholarly, trade, or popular, but as there is no official list assigning publishers to these categories, this was not feasible.

The third independent variable, platform, encompasses several differences between platforms. Platforms differ in how many pages from each book can be printed, software

requirements for downloading, the quality of the bibliographic records they provide, and how books are exposed in Google, for example. Some of these features can also differ within a platform. An initial attempt was made to compare indexing in Google as it seemed likely that the level of indexing in Google would affect whether a book was used. Students and faculty are more likely to discover e-books through general internet searches than through the library catalog.[63] SpringerLink has noted that half of all traffic to their site is from search engines and only 20 percent from library tools.[64] Discoverability proved to be difficult to measure, since information on indexing could only be found through personal contacts with vendors who did not provide information in a standardized way. In the end, platform was used as a variable with the understanding that platforms differ, and an observed difference in usage between platforms should not be attributed to any particular features of that platform.

To make comparisons with the print collection, an additional sample was taken of print books. Like the e-books, this sample was limited to books published in 2015. The list was compiled using a report from the library catalog, limited to books held by the main library and published in 2015. After exporting the list to Excel, a column was added that extracted the first letter of each book's call number so that print books could be categorized by single-letter LC class, as was done for the e-books.

To determine the extent of usage for a certain subset of books, PEU was calculated for print and e-books. Calculations were based solely on the sample, not the full collection, and were done separately for each format. PEU was calculated as the percentage of all books used from this category and was divided by the percentage of all available books that were in the category. For instance, 1.65 percent of books in the print book sample were in LC class F, while 2.08 percent of the print books that were used in 2016 were in LC class F. The percent of used titles divided by the percent of available titles (2.08 ÷ 1.65) yields a PEU of 1.26.

The dependent variable in the study was usage in 2016. As mentioned above, BR2 tracks the number of sections that have been viewed in each book, but the definition of a section varies by platform. Several of the vendors in this study counted each page viewed as one section, while others counted each chapter. The measure that could be compared across platforms, first suggested by Littman and Connaway and later supported by Knowlton, was a simple yes/no count of whether a title was used.[65] The same measure was used for print books.

Once all the variables were calculated, several comparisons were done using logistic regression, a statistical method that produces an equation that calculates the log of the odds of a specific outcome. In this case, the outcome is expected use of a book. A higher log odds means that the book is more likely to be used. A regression equation can contain several independent variables, or predictor variables, which are correlated with higher or lower odds of the desired outcome occurring. The goal was to see which variables had the strongest correlation with the desired outcome, i.e., e-books being used. This paper focuses on whether a particular feature of an e-book increases or decreases the odds of it being used rather than calculating the actual odds.

Data

Before putting any of the variables into a regression equation, crosstabs were used to explore each variable separately to identify which variables appeared to be related to differences in e-book use. Table 1 shows that there are differences between the five platforms in what percentage of the books available on that platform were used in 2016. The p-value underneath the table (p < .001) indicates that it is statistically highly unlikely that there would be no significant difference between the full e-book collections given what was observed in the sample.

The next variable considered was LC class. For this, e-books from all five platforms were grouped together and comparisons were made across LC main class. Table 2 shows that, across all platforms, there are differences between subjects regarding how many e-books are used. The classes with the highest percentage of books used are A (General Works) and Z (Bibliographies and Library Science), followed by R (Medicine). The V section (Naval Science) has the smallest percentage of books used, at only 5.26 percent, but since the sample contains only nineteen books in this section, this is not an area of focus for this library. As before, the p-value listed below the table indicates that it is highly unlikely that there would be no differences in usage between the classes in the full collections from which the sample is drawn.

The third variable considered was publisher type, which was coded as university press or other. In table 3, it is clear that books from non-university press publishers are used much more than university press books. Again, there is a high level of statistical significance, i.e., a low p-value.

Once each of the variables was individually examined, and analysis had shown that there are differences in usage depending on a book's platform, LC class, and publisher type, the variables were placed in a logistic regression model. At this stage, certain LC classes were removed. Call numbers beginning with K were removed, as print books in this area are held in a separate law library, so there would not be print data to compare with these e-books. Class V was removed as only one of these e-books was used. A forward-selection modeling technique was used, meaning the initial regression equation used only one independent

Table 1. Differences in Usage by Platform

Provider	% Used	*n*
ebrary	18.20	10,368
MyiLibrary	15.18	4,314
netBASE	8.53	434
Springer	23.03	6,856
Wiley	13.84	2,450

Chi-sq = 189.9862, df = 4, p < .001

Table 2. Differences in Usage by LC Class

LC Class	% Used	*n*
A–General Works	34.62	26
B–Philosophy, Psychology, Religion	13.72	1,713
C–History	23.30	103
D–World History	13.40	933
E–History of the Americas	10.35	425
F–History of the Americas	12.65	332
G–Geography, Anthropology, Recreation	20.93	688
H–Social Sciences	17.17	5,184
J–Political Science	13.62	727
K–Law	21.07	598
L–Education	24.25	1,068
M–Music	24.66	219
N–Fine Arts	26.91	405
P–Language and Literature	13.47	2,338
Q–Science	19.39	4,022
R–Medicine	30.88	2,273
S–Agriculture	16.94	301
T–Technology	15.22	2,812
U–Military Science	14.29	91
V–Naval Sciences	5.26	19
Z–Bibliography, Library Science	32.41	145

Chi-square = 461.7503, df = 20, p < .001

variable and then another variable was added to create model 2 and then another in model 3. The goal is to obtain a model where all the variables show some degree of statistically significant correlation with the outcome.

The numbers shown in table 4 are coefficients that would be used to create a regression equation. Each coefficient is listed, along with its accompanying standard error. The coefficients indicate how much the log odds of an e-book being used will be affected by the variable in question. When the variables in a regression equation are categories rather than numbers (e.g., LC class rather than

Table 3. Differences in Usage by Publisher Type

	% Used	*n*
University press	8.22	4,583
Other	20.77	19,839

Z = 19.7425, p < 0.001

year of publication, for example), one of the categories is always treated as a reference category. In table 4, the reference category for LC class is class A, which means that there is no coefficient listed for class A in table 4. Rather, all the other LC classes are considered in terms of whether or not they are more likely to be used than class A. A negative coefficient means books in this class have a lower likelihood of being used than books in class A. For instance, if the variable class L has a coefficient of -0.5, then for e-books in class L the log of the odds of their being used will be .5 lower than the log odds of the reference group (class A) if all other variables are held constant. To find the actual odds, take the anti-log of the log odds.

Model 1, shown in table 4, examines only the usefulness of the LC class in predicting the likelihood of an e-book being used. Statistically significant relationships are marked with asterisks indicating the p-value. A value with no asterisk represents a finding that is not statistically significant, that is, the p-value is above .05. Lack of statistical significance means it is possible that the difference in usage between these books and others in the sample would not hold true in the full collections of e-books. For the classes with statistical significance, the number in the table indicates how much the log odds of the book being used will be affected by the book being in that class.

Model 2 introduces the university press variable and model 3 introduces platform. Platforms were introduced last in the model as they serve as a catch-all, representing several other unmeasurable differences between the books, such as interface design and discoverability via Google.

The last row in the table, McFadden's Pseudo-R^2, is a goodness-of-fit measure that tells how much of the variation in usage rates can be explained by the predictors included in the regression equation. Model 3 has a pseudo-R^2 of 0.0396, indicating that 3.96 percent of the variation can be explained by the variables in the model. Since model 3 includes the largest number of variables with significant correlations, and has the largest pseudo-R^2, it has the most explanatory power. Because the table shows three different models, each time a variable is added to the model, it increases the model's explanatory power. When university press was added in model 2, the pseudo-R^2 increased from 0.0192 to 0.0338, while adding platform in model 3 increased it only slightly to 0.0396. It seems that in this dataset, university press status adds the most explanatory power.

Table 4. Logistic Regression Models

Variables	Coefficients					
	Model 1		Model 2		Model 3	
	coefficient	s.e.	coefficient	s.e.	coefficient	s.e.
(Intercept)	-.64	0.41	-0.4	0.42	-0.35	0.42
LC class (*reference A*)						
B	-1.20**	0.42	-1.23**	0.43	-1.21**	0.43
C	0.56	0.47	-0.52	0.48	-0.53	0.48
D	-1.23**	0.42	-1.11*	0.43	-1.09*	0.43
E	-1.52***	0.44	-1.12*	0.45	-1.09*	0.45
F	-1.30**	0.44	-0.89*	0.45	-0.90*	0.45
G	-0.69	0.42	-0.73	0.43	-0.71	0.45
H	-0.94*	0.41	-1.07*	0.42	-0.95*	0.43
J	-1.21**	0.42	-1.18**	0.43	-1.15**	0.43
L	-0.50	0.42	-0.67	0.46	-0.66	0.43
M	-0.48	0.44	-0.29	0.42	-0.26	0.45
N	-0.36	0.43	-0.41	0.45	-0.39	0.44
P	-1.22**	0.42	-1.12**	0.44	-1.11**	0.43
Q	-0.79	0.42	-0.97*	0.42	-1.00*	0.42
R	-0.17	0.41	-0.36	0.42	-0.37	0.42
T	-1.08**	0.42	-1.29**	0.42	-1.22**	0.42
Z	-0.10	0.45	-0.19	0.46	-0.18	0.46
University Press	-		-1.02***	0.06	-1.04***	0.06
Provider (*reference ebrary*)						
MyiLibrary	-		-		-0.22***	0.05
netBASE	-		-		-1.05***	0.19
Springer	-		-		0.05	0.05
Wiley	-		-		-0.52***	0.07
McFadden's Pseudo-R²	.0192		.0338		.0396	

* p < .05
**p < .01
***p < .001

All the platforms, except for Springer, have statistically significant correlations. Although the sample shows that Springer e-books have a higher rate of usage than Ebrary books, it is somewhat likely that these differences are a chance outcome due to content in this particular sample. The remaining three platforms show statistically significant correlations, and they are all significantly less likely to be used than the Ebrary e-books, with other variables held constant.

After using logistic regression to identify which specific subjects are more likely to be used, a secondary question arose of whether it is possible to generalize about which subjects get more use. Specifically, do subjects with heavy print use receive less e-book use? To answer this question, the print PEU for each class was compared to the e-book PEU.

Although the study's overall intent was to combine multiple variables into the same regression model, this second question required a separate analysis. Since PEU is calculated for each LC class rather than each book, the PEU for a given LC class is always the same. A book with LC class G will always have a Print PEU of 1.31, and every book with LC class H will have a Print PEU of .99. A regression equation cannot contain two independent variables whose values correspond perfectly. Therefore, the relationship between print and e-book use was examined in a separate analysis.

The first step was to create a scatterplot, shown in figure 1, to see if there appeared to be a relationship between the PEU of a certain class of books in print and the same class of books in e-book form. In figure 1, each dot

represents an LC class. If books with a higher print PEU consistently had a lower e-book PEU, the dots would arrange themselves in a line sloping from the top left of the plot (high e-book PEU, low print PEU) to the bottom right (low e-book PEU, high print PEU). In fact, there does not appear to be a relationship, and this was confirmed by statistical analysis. A linear regression equation that attempted to find a correlation between print PEU and e-book PEU returned a p-value of 0.9822, indicating that there is not a statistically significant relationship between these two variables. This is in contrast to Slater's finding of a positive correlation.[66]
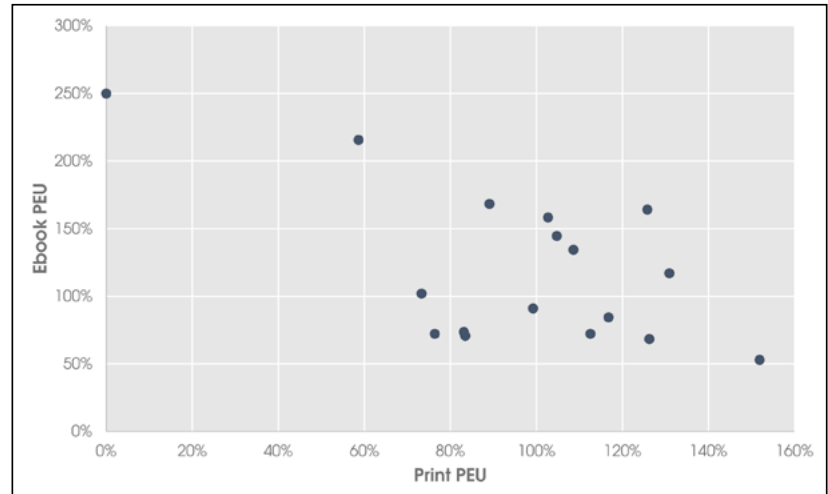

Figure 1. Print Usage Plotted Against E-book Usage

Discussion

By using logistic regression, this study identified what factors are most useful in predicting which e-books will be used. Those most likely to be used are e-books in the LC class A, published by non-university presses, and hosted on Ebrary. However, a substantial amount of the variation in use between different e-books is not explained by the regression equation provided here and is due to an unknown factor.

The finding that Ebrary books receive the most use is surprising since they are part of a subscription collection. Recall Levine-Clark's observation that Ebrary books were used less than EBL, and his hypothesis that selection method (title-by-title versus subscription package) accounted for the low use in Ebrary.[67] Though EBL is not included in this study, the data here includes MyiLibrary, which contains a combination of DDA and firm order titles. One might expect these to be used more often than the Ebrary package, but that is not the case. Librarians at this institution indicated a strong preference for Ebrary's interface over MyiLibrary, lending support to the interpretation that usability affects the likelihood of an e-book being used. Since both platforms have migrated to ProQuest's Ebook Central since the time of this analysis, it is possible that usage of the books formerly hosted on MyiLibrary will increase. Another factor that could possibly explain the differing usage is the amount of detail included in catalog records, a variable that was not examined in this study. This would differ by title, but verification of randomly selected records from the Ebrary and MyiLibrary datasets found that content notes appear to be more common in records for Ebrary books than for MyiLibrary.

The finding that university press books were used less than other books, when platform and subject are held constant, was less of a surprise. In Levine-Clark, Paulson, and Moeller's study they distinguish university press books from others because they see this as a proxy for a book being of especially high quality and they find these to have higher use.[68] However, Christianson and Aucoin had the opposite finding, and this study corroborates theirs in that way.[69] The presumed explanation in this study is similar to that of Christianson and Aucoin. Since university press books are often on narrow topics, it is expected that they would appeal to fewer users despite their high quality. University press books would be more likely to be used by faculty or graduate students who together make up only half as much of Temple University's population as undergraduates.

The LC class with the highest rate of usage is A, General Works. This is unexpected as general works are inherently not an area of focus. A look at the titles used shows that some are related to digital humanities, which is an area of focus for the main library. It is not surprising that math and science books (class Q) had one of the highest odds of being used, with other variables held constant, as this was noted in several other studies.[70] This could be because science books are less likely to be intended for linear reading. Among the higher rates of usage is history of the Americas (class F). History is traditionally a discipline that has expressed a preference for print, likely due to the nature of history materials, which typically involve narrative. The PEU calculations corroborate this preference, demonstrating that even when there is a strong preference for print, e-books can still receive some use. Technology books (class T) had the lowest odds of e-book use despite the more common finding of this being a popular topic for e-books.[71] Technology actually showed a strong preference for print, with a print PEU of 1.17 and an e-book PEU of only .84. Although this contradicts the usual assumption of technology being a popular topic for e-books, the LC class includes photography

books, which are preferred in print due to image quality. The low usage rates may also be because the library has several technology-focused databases, and users may prefer these to an aggregator package such as Ebrary.

While the findings for specific subjects are relevant to selectors, the more substantial finding is that other e-book features have a stronger correlation with usage than does the subject matter. Furthermore, the regression equation underscores the fact that most of the factors influencing usage have not yet been identified. The variables included in this study only predict 3.96 percent of the variation in e-book use. Future studies might be able to use the methodology presented here as a model for exploring the effects of additional variables on e-book use. Such studies should also attempt to consider specific platform characteristics rather than measuring platform as a single variable. It is likely that some of the differences in use rates between platforms are due to platform characteristics that were not examined in this study, such as ease of finding books using Google, the quality of the bibliographic records in the catalog, the reputation of the publishers represented in the collection, or the selection method for books on a given platform. The finding that Ebrary books have the highest use rate does not necessary mean libraries should acquire books only from this provider (now Ebook Central). If the difference in usage between platforms can be shown to correspond with particular interface features, the library could pursue purchases on new platforms with interfaces that are equally good. If the difference is due to discoverability in Google, then the library could make that a priority in selecting platforms. Ideally, further research would incorporate additional features of e-books and would be able to separate the effect of these features from unidentified other differences in platforms.

Conclusion

The model offered here can contribute to the body of literature that is gradually accumulating showing how e-book use differs by subject and provider. More importantly, it provides an example of one way to tease out the variety of factors that influence e-book usage. In response to Wilkin and Underwood's statement that "researchers are interpreting the issue of what constitutes the 'e-book problem' differently," this research suggests a way to unify the various research questions of previous studies into one overarching question: what factors predict e-book use?[72] Though this study considers just three variables, it offers a methodology that can incorporate further variables.

In addition to providing a unified research question, this paper contributes toward building a standard for measuring e-book use by relying on emerging conventions. The literature shows that some standard methods are beginning to emerge. Comparing books based on whether they receive use rather than the amount of usage is a method that will hopefully become standard. The PEU as a unit of comparison is a well-established measure that can be used for both print and e-books. The means of finding an appropriate print collection to compare to the e-books under consideration will vary depending on the library's holdings. Knowlton and Fry offer methods that could work for any institution.[73] The tactic for finding a call number does not need to be consistent across studies, though when multiple studies use LC call numbers rather than vendor-provided subject categories, it is easier to compare them to each other.

By using the yes/no measure of use, PEU, and LC classes, this paper presents findings in a way that they can be compared against other studies to build a broad sense of e-book use in academic libraries. It would be very helpful to see future research that also takes into account whether a title was selected by a librarian, as part of a package, through an approval plan, or as a patron-driven acquisition. A more granular analysis of publisher types would also be helpful. Despite these gaps, a large enough body of work is emerging that results can be aggregated to provide some answer to the general question of which e-books get used. Though e-book usage may still present what Wilkin and Underwood call a "wicked problem," librarians are gradually working their way toward standards of measurement that will allow not only for more analysis at the institutional level but for comparisons between different studies that will produce better informed decisions

References

1. James Cory Tucker, "Ebook Collection Analysis: Subject and Publisher Trends," *Collection Building* 31, no. 2 (2012): 40–47.
2. Robert Slater, "E-Books or Print Books, 'Big Deals' Or Local Selections—What Gets More Use?," *Library Collections, Acquisitions & Technical Services* 33, no. 1 (2009): 40.
3. Timothy P. Bailey, "Electronic Book Usage at a Master's Level I University: A Longitudinal Study," *Journal of*

Academic Librarianship 32, no. 1 (2006): 52–59; Steven B. Carrico et al., "What Cost and Usage Data Reveals about E-Book Acquisitions: Ramifications for Collection Development," *Library Resources & Technical Services* 59, no. 3 (2015): 102–11; Marilyn Christianson and Marsha Aucoin, "Electronic Or Print Books: Which are Used?," *Library Collections, Acquisitions & Technical Services* 29, no. 1 (2005): 71–81; Amy Fry, "Factors Affecting the Use of Print and

Electronic Books: A Use Study and Discussion," *College & Research Libraries* 79, no. 1 (2018): 68–85; Steven A. Knowlton, "A Two-Step Model for Assessing Relative Interest in E-Books Compared to Print," *College & Research Libraries* 77, no. 1 (2016): 20–33; Nancy Sprague and Ben Hunter, "Assessing E-books: Taking a Closer Look at E-book Statistics," *Library Collections, Acquisitions & Technical Services* 32, no. 3 (2008): 150–57; Tucker, "Ebook Collection Analysis."

4. Shelley Wilkin and Peter G. Underwood, "Research on E-book Usage in Academic Libraries: 'Tame' Solution or a 'Wicked Problem'?," *South African Journal of Libraries & Information Science* 81, no. 2 (2015): 11–18.

5. Knowlton, "A Two-Step Model"; Fry, "Factors Affecting the Use of Print and Electronic Books."

6. Carrico et al., "What Cost and Usage Data Reveals," 102–11.

7. Michael Levine-Clark, "Global Trends in Ebook Usage: Patterns from 10,000 Libraries," ProQuest, 2015, http://contentz.mkt5049.com/lp/43888/413459/2015_Ebooks-Usage-Data-White-Paper_MLC.pdf.

8. Terry Bucknell, "The Big Deal Approach to Acquiring eBooks: A Usage-Based Study," *Serials* 23, no. 2 (2010): 126–34; Alain R. Lamothe, "Factors Influencing the Usage of an Electronic Book Collection: Size of the E-book Collection, the Student Population, and the Faculty Population," *College & Research Libraries* 74, no. 1 (2013): 39–50.

9. Bucknell, "The Big Deal Approach," 126, 134; Matthew Connor Sullivan and Katherine Leach, "Hard Data for Tough Choices: eBooks and pBooks in Academic Libraries" (presentation, E&RL Conference, Austin, Texas, April 5, 2016), http://schd.ws/hosted_files/erl2016/51/ER%26L%202016_Hard%20data%20for%20tough%20choices.pptx.

10. Michael Levine-Clark, Kari Paulson, and Paul Moeller, "10,000 Libraries, 4 Years: A Large-Scale Study of E-book Usage and How You Can Use the Data to Move Forward," *Serials Librarian* 68, no. 1–4 (2015): 262–68; Christianson and Aucoin, "Electronic or Print Books"; Tucker, "Ebook Collection Analysis."

11. Tucker, "Ebook Collection Analysis."

12. Slater, "E-Books or Print Books"; Carrico et al., "What Cost and Usage Data Reveals."

13. Fry, "Factors Affecting the Use of Print and Electronic Books"; Wilkin and Underwood, "Research on E-Book Usage," 15.

14. Levine-Clark et al., "10,000 Libraries, 4 Years."

15. Wilkin and Underwood, "Research on E-Book Usage."

16. Project COUNTER, "The COUNTER Code of Practice for E-resources: Release 4," April 2012, www.projectcounter.org/wp-content/uploads/2016/01/COPR4.pdf.

17. Project COUNTER, "The COUNTER Code of Practice," (2017), accessed August 30, 2017, www.projectcounter.org/code-of-practice-sections/general-information/.

18. Karin Bystrom, "Everything that's Wrong with E-book Statistics: A Comparison of E-book Packages" (presentation, 32nd Annual Charleston Conference Issues in Book and Serial Acquisition, Charleston, South Carolina, November 9, 2012).

19. John Cox, "Making Sense of E-book Usage Data," *Acquisitions Librarian* 19, no. 3–4 (2008): 193–212.

20. Bucknell, "The Big Deal Approach," 126–34.

21. Lamothe, "Factors Influencing the Usage"; Tucker, "Ebook Collection Analysis."

22. Levine-Clark et al., "10,000 Libraries, 4 Years."

23. Michael Levine-Clark, "E-book Usage on a Global Scale: Patterns, Trends and Opportunities," *Insights: The UKSG Journal* 28, no. 2 (2015): 39–48.

24. Justin Littman and Lynn Silipigni Connaway, "A Circulation Analysis of Print Books and E-Books in an Academic Research Library," *Library Resources & Technical Services* 48, no. 4 (2004): 256–62; Lisa Rose-Wiles, "Are Print Books Dead? An Investigation of Book Circulation at a Mid-Sized Academic Library," *Technical Services Quarterly* 30, no. 2 (2013): 129–52; Slater, "E-books or Print Books"; Sullivan and Leach, "Hard Data for Tough Choices;" Tucker, "Ebook Collection Analysis."

25. Knowlton, "A Two-Step Model."

26. Ibid., 21.

27. Rusty Kimball, Gary Ives, and Kathy Jackson, "Comparative Usage of Science E-book and Print Collections at Texas A&M University Libraries," *Collection Management* 35, no. 1 (2009): 15–28.

28. Ibid.

29. Knowlton, "A Two-Step Model."

30. Christianson and Aucoin, "Electronic or Print Books"; Cathy Goodwin, "The E-Duke Scholarly Collection: E-book v. Print Use," *Collection Building* 33, no. 4 (2014): 101–5; Littman and Connaway, "A Circulation Analysis of Print Books and E-books"; Slater, "E-books or Print Books"; Sullivan and Leach, "Hard Data for Tough Choices."

31. Goodwin, "The E-Duke Scholarly Collection."

32. Kimball et al., "Comparative Usage of Science E-book and Print Collections"; Littman and Connaway, "A Circulation Analysis of Print Books and E-books"; Slater, "E-books or Print Books."

33. Fry, "Factors Affecting the Use of Print and Electronic Books."

34. Knowlton, "A Two-Step Model."

35. Fry, "Factors Affecting the Use of Print and Electronic Books."

36. Knowlton, "A Two-Step Model."

37. Ibid.

38. Slater, "E-books or Print Books."

39. Bailey, "Electronic Book Usage at a Master's Level I University"; Christianson and Aucoin, "Electronic or Print Books."

40. Tucker, "Ebook Collection Analysis."

41. Carrico et al, "What Cost and Usage Data Reveals."

42. Littman and Connaway, "A Circulation Analysis of Print Books and E-books."

43. Knowlton, "A Two-Step Model"; Fry, "Factors Affecting the Use of Print and Electronic Books"; Sprague and Hunter "Assessing E-books."

44. Knowlton, "A Two-Step Model"; Fry, "Factors Affecting the Use of Print and Electronic Books"; Sprague and Hunter, "Assessing E-Books"; Tucker, "Ebook Collection Analysis."

45. Goodwin, "The E-Duke Scholarly Collection."

46. Slater, "E-books or Print Books."

47. Knowlton, "A Two-Step Model."

48. Sprague and Hunter, "Assessing E-books."

49. Knowlton, "A Two-Step Model."

50. Slater, "E-books or Print Books."

51. Littman and Connaway, "A Circulation Analysis of Print Books and E-books."

52. Christianson and Aucoin, "Electronic or Print Books."

53. Slater, "E-books or Print Books."

54. Christianson and Aucoin, "Electronic or Print Books."

55. Littman and Connaway, "A Circulation Analysis of Print Books and E-books."

56. Sullivan and Leach, "Hard Data for Tough Choices."

57. Kendall Hobbs and Diane Klare, "Are We There Yet? A Longitudinal Look at E-books through Students' Eyes," *Journal of Electronic Resources Librarianship* 28, no. 1 (2016): 9–24.

58. Littman and Connaway, "A Circulation Analysis of Print Books and E-books."

59. Christianson and Aucoin, "Electronic or Print Books"; Sullivan and Leach, "Hard Data for Tough Choices."

60. Christianson and Aucoin, "Electronic or Print Books."

61. Levine-Clark et al., "10,000 Libraries, 4 Years."

62. Hobbs and Klare, "Are We There Yet?"; Slater, "Why Aren't E-books Gaining More Ground in Academic Libraries?"

63. Anna Faherty, "Academic Book Discovery, Evaluation and Access: Insights and Opportunities for Enhancing the Scholarly Experience," June 2016, https://academicbookfuture.files.wordpress.com/2016/06/faherty_academic-book-discovery-full-report.pdf; Hobbs and Klare, "Are We There Yet?"

64. Timon Oefelein, "Global Discovery Trends and the Library's Changing Role," *Research Information*, May 5, 2016, www.researchinformation.info/news/analysis-opinion/global-discovery-trends-and-librarys-changing-role.

65. Knowlton, "A Two-Step Model"; Littman and Connaway, "A Circulation Analysis of Print Books and E-books."

66. Slater, "E-books or Print Books."

67. Levine-Clark, "Global Trends in Ebook Usage."

68. Levine-Clark et al., "10,000 Libraries, 4 Years."

69. Christianson and Aucoin, "Electronic or Print Books."

70. Slater, "E-books or Print Books"; Sprague and Hunter "Assessing E-books."

71. Slater, "E-books or Print Books."

72. Wilkin and Underwood, "Research on E-book Usage," 14.

73. Fry, "Factors Affecting the Use of Print and Electronic Books"; Knowlton, "A Two-Step Model."