7-2015

# E-Book Usage on a Global Scale: Patterns, Trends, and Opportunities

Michael Levine-Clark
*University of Denver*

# E-Book Usage on a Global Scale: Patterns, Trends, and Opportunities

## Comments

Published in *Insights.* May be found online: https://doi.org/10.1629/uksg.240

## Publication Statement

## Publication Statement

UKSG

# E-book usage on a global scale: patterns, trends and opportunities

*Based on a breakout session presented at the 38th UKSG Annual Conference, Glasgow, March 2015*

This study examines worldwide usage of over 600,000 e-books from Ebook Library (EBL) and ebrary. Using multiple modes of analysis, the study shows that there are variations in usage by geographic region as well as by subject. The study examines usage in relation to availability of titles, different types of usage per session, usage of the top ten percent of titles, and intensive and extensive use. These patterns can be used for benchmarking and as a model for local e-book studies.

## Introduction

As ProQuest began planning the future integration of Ebook Library (EBL) and ebrary, they decided to look at past usage of e-books on both platforms. They commissioned a white paper on worldwide usage of titles from ebrary and EBL as a way of better understanding how libraries have used e-books so far. The white paper had not been published at the time of the UKSG conference, but will be available by the time this summary article is published, and will provide greater detail[1]. After presenting at several conferences in the US over the past year, at UKSG the author and Kari Paulson presented the most comprehensive overview of the research to date[2]. Consisting of over 640,000 titles on the ebrary platform and around 375,000 on EBL, this is a massive data set that allows for a greater understanding of usage patterns on a broad scale.

MICHAEL
LEVINE-CLARK

Associate Dean
for Scholarly
Communication and
Collections Services
University of Denver
USA

While a data set of this size can certainly provide generalizable insights about usage, it is important to acknowledge that local patterns may vary significantly. Each institution has a unique curricular and research focus, which drives both collection development and usage. A university with a strong focus on the arts will see very different levels of usage from one with a focus on engineering. And neither of them may have the same types of use as the libraries in general across the world. In addition, the other materials available at any given institution will have an impact on the usage of the specific titles available from ebrary and EBL. A library with many collections of e-books, each competing for the user's attention, may see generally lower levels of usage of ebrary or EBL books than a library with only ebrary or EBL collections. A library with a large collection of science e-books on publisher platforms may also see lower usage of science titles on the ebrary or EBL platform. While these local differences are important, the general patterns shown in the data can be useful for understanding broad disciplinary differences and for benchmarking local patterns against expected norms.

> 'a massive data set that allows for a greater understanding of usage patterns on a broad scale'

## The data set

This study examines usage of ebrary titles in 2014 and EBL titles in 2013. Both sets include all customers worldwide, and account for multiple types of usage. EBL and ebrary have not yet merged their data, so this study looks at them as separate data sets. There are some measures of use only included for one and not the other. For example, while we have a regional breakdown of usage for EBL, we do not have a regional breakdown of the number

of titles available. We have both types of data for ebrary, and therefore focus on ebrary when making some regional comparisons.

For the most part, the data have commonalities in terms of what is included. There are multiple usage metrics that together allow for an expansive picture of usage. At the broadest level is a session, which is the most basic level of interaction with an e-book. Within a session, one or more of the other types of usage can occur. These include a count of:

- the number of pages viewed, pages copied/pasted, or pages printed

- usage time

- downloads, either of the book or the chapter.

While usage time and pages viewed are a measure of known use, downloading, printing and copying allow for some usage that is not included in the data. For instance, a download of an entire book may lead to one page being read, the entire book being read, or a PDF sitting unread on a patron's computer.

In addition to a variety of measures of usage, the data set includes metrics about the number of libraries with the title visible and the number of libraries with usage, aggregated at the title level. The call numbers for each title were used to look at usage at the Library of Congress (LC) Class and Subclass level and were also used to build broader sets of titles at the level of academic division (arts and humanities; social sciences; and science, technology, engineering and medicine [STEM]). See Table 1 for an overview of these academic divisions and the LC Classes and Subclasses used in their construction.

| Arts and humanities | Social sciences |
| --- | --- |
| B-BD, BH-BX (Philosophy, Religion) | BF (Psychology) |
| C, D, E, F (History) | H (Social Sciences) |
| M (Music) | J (Political Science) |
| N (Fine Arts) | K (Law) |
| P (Language & Literature) | L (Education) |
| TR (Photography) | U (Military Science) |
| | V (Naval Science) |
| **STEM** | **Other** |
| Q (Science) | A (General Works) |
| R (Medicine) | G (Geography, Anthropology, Recreation) |
| S (Agriculture) | Z (Bibliography, Library Science, Information Resources) |
| T-TP, TS-TX (Technology) | |

Table 1. Academic Divisions with Library of Congress Classes and Subclasses

Across the entire data set for ebrary, there are 642,885 titles that were visible to at least one library in 2014. North American libraries had almost as many titles available, followed by European libraries with under 500,000 visible to their users. See Figure 1 for an overview of the number of titles visible to libraries by region. A large portion of the titles visible worldwide from ebrary were in a small portion of LC Classes. The H (Social Sciences) class, which includes Business, accounted for 18.5% of all titles available, with a small number of other subjects accounting for a large percentage of available titles. The patterns are similar for EBL. See Figures 2 and 3 for title visibility by LC Class for ebrary and EBL respectively. At a broader disciplinary level, for ebrary, STEM subjects accounted for 30.2% of the titles available, arts and humanities accounted for 28.9%, and social sciences broadly (including more than just the LC Class H, as shown in Figure 1) accounted for 28.6%. For EBL, social sciences accounted for the most titles (31.3%), followed by STEM (29.5%), and arts and humanities (26.6%).
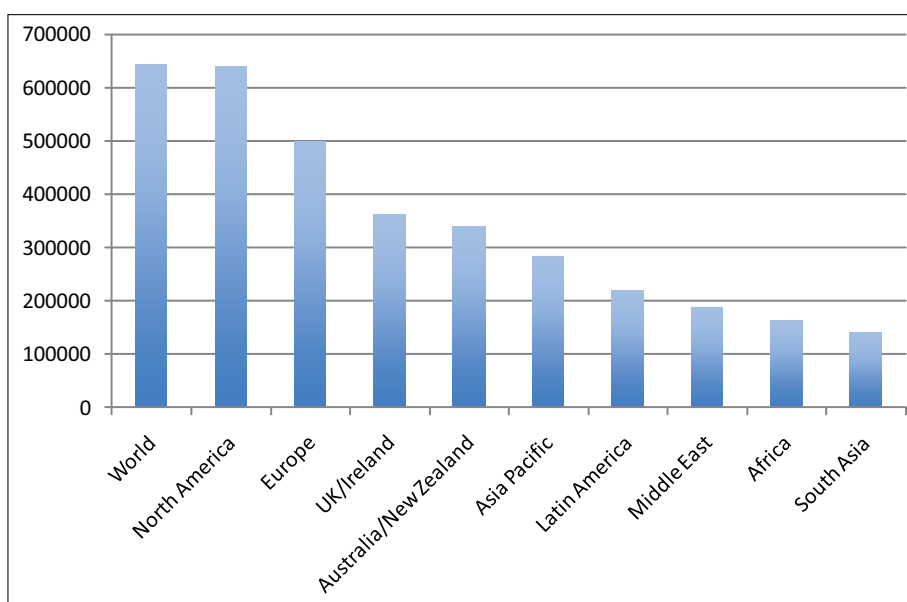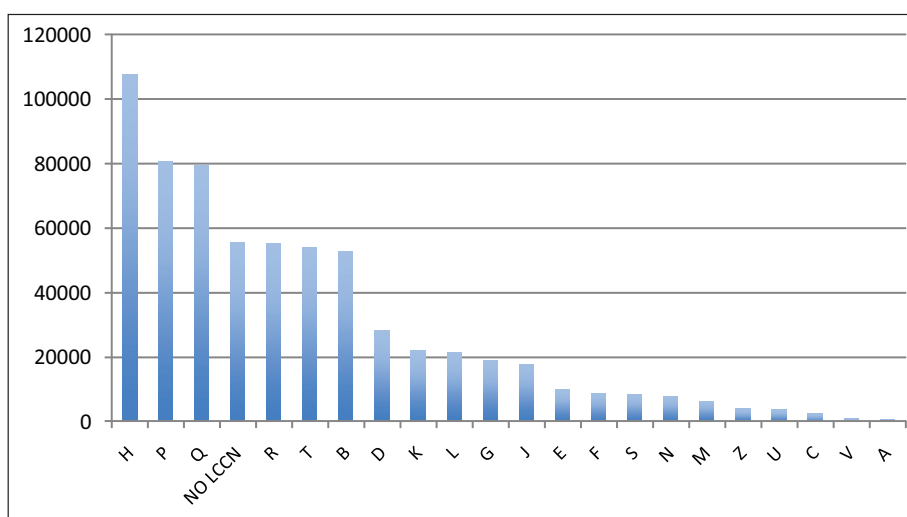
Figure 1. Titles available: ebrary



Figure 2. Titles available worldwide, by LC Class: ebrary
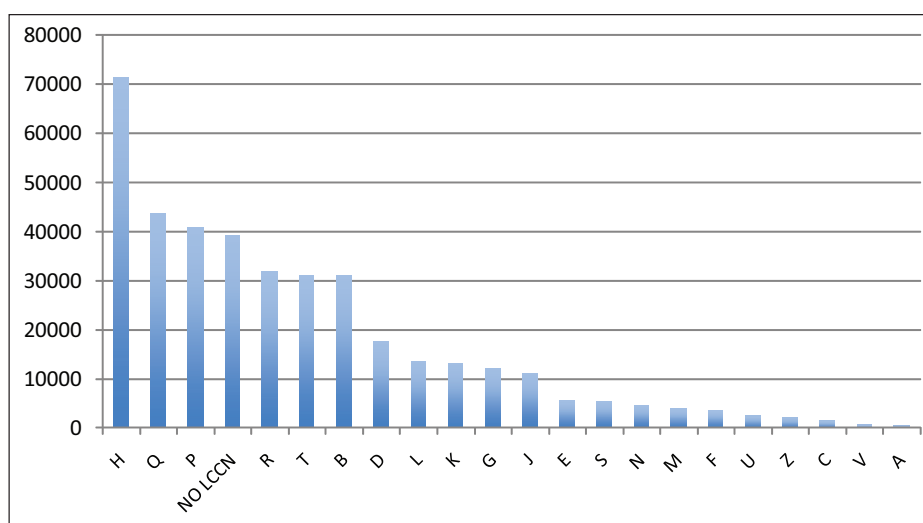


Figure 3. Titles available worldwide, by LC Class: EBL

The level of visibility of titles by library varies. On average for ebrary, each title was available in 412 libraries, with arts and humanities titles being available in 555 libraries on average, social sciences in 509 and STEM titles in 379. For EBL, the numbers are lower overall,

with titles visible on average in 89 libraries. The disciplinary breakdowns are different for EBL, with just under 101 libraries having access to an average title in the social sciences, 88 libraries having access to an average STEM title, and only 77 having access to a title in the arts and humanities. Some of this variability can be explained by 'Academic Complete', the ebrary subscription package of around 125,000 titles, available in its entirety in about 2,000 libraries worldwide. Academic Complete is most heavily weighted toward the arts and humanities (38.0%), then social sciences (32.4%), followed by STEM (25.7%), which matches the visibility patterns overall for ebrary titles.

## Categories of usage

There are multiple ways of considering use. The following sections explore four broad categories of usage:

1. Usage compared to availability: the percentage of titles visible across the entire data set that are used at any library

2. Usage patterns per session: the level of usage of different types for every session that occurs

3. Usage of the most popular titles: the level of usage for the titles with the top 10% of sessions

4. Intensive vs. extensive use: a comparison of the percentage of titles in a given LC Class that are used with the number of times they are used.

Together, these four categories provide a nuanced measure of usage that may help us better understand usage overall. (These and other categories are explored in more depth in the white paper.)

### Usage compared to availability

Studying usage relative to availability is helpful for understanding whether it makes sense for a library to invest in a given subject and takes into account the fact that there are large differences in the number of titles available across subjects. While there are far more titles with at least one session in the Hs (Social Science) than in the Es (History of the Americas), there are also far more titles available in the Hs. Looking at the percentage of titles used in each subject area shows that Es are most likely to be used for ebrary, with 73.3% of titles in this subject area having at least one session worldwide compared to 64.1% of the Hs. See Figures 4 and 5 for a breakdown of usage compared to availability by LC Class for ebrary.

> 'Studying usage relative to availability is helpful for understanding whether it makes sense for a library to invest'
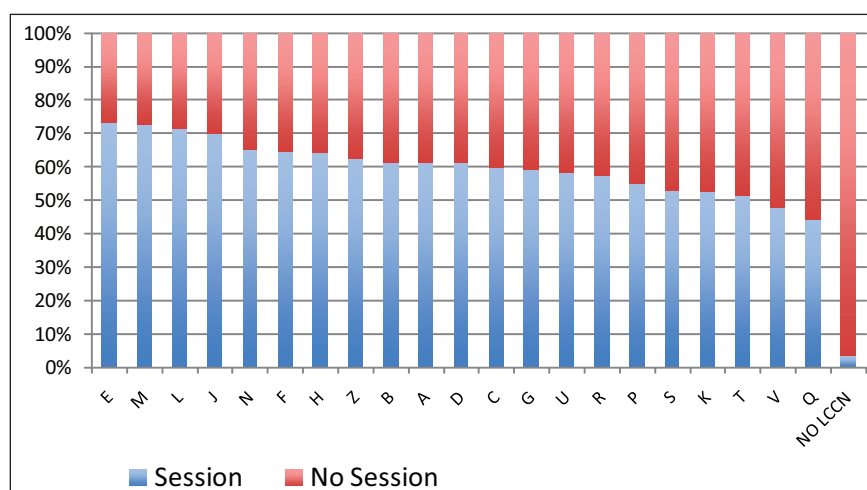


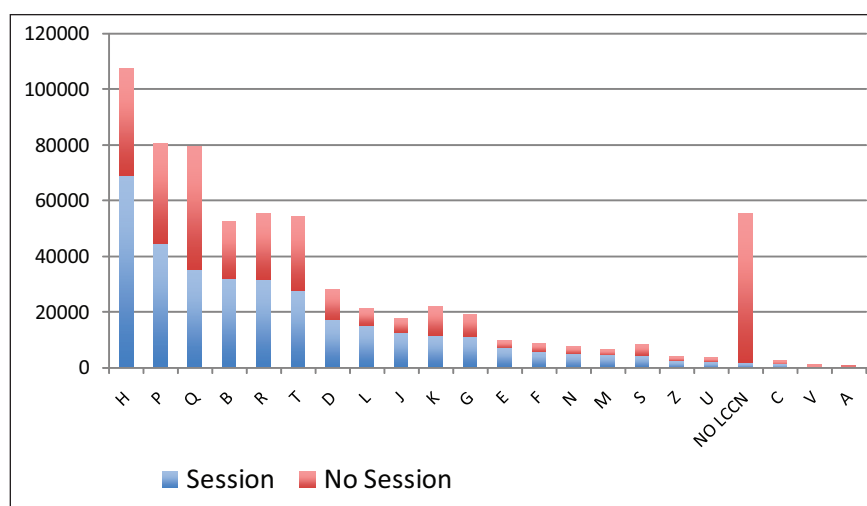Figure 4. Percentage of titles used, by LC Class: ebrary

Figure 5. Titles used, by LC Class: ebrary

Stepping back from individual subjects, just over 53% of all titles that were available on the ebrary platform in 2014 had a session, with the percentage of titles used going down from there to 25.4% for Africa (see Figure 6). For EBL, almost 69% of titles were used in 2013. This disparity is connected to the large Academic Complete subscription package; it is likely that ebrary customers, on average, have more titles available to them and therefore will see a lower percentage of use. It is worth noting that for each platform this is usage over a single-year period, and multi-year usage would almost certainly show a higher percentage of titles used overall. As Figures 7 and 8 show, the percentage of titles used by discipline varies between ebrary and EBL, though in both cases social sciences titles are most likely to be used. The higher usage rate for STEM titles on the EBL platform matches the higher overall availability across libraries for those disciplines on that platform.
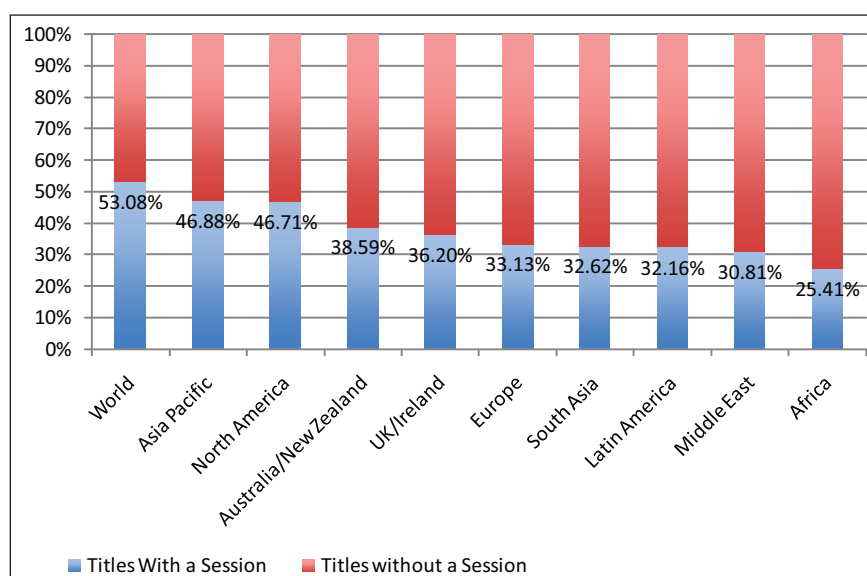


Figure 6. Titles available / Titles used: ebrary

Another way of considering usage compared to availability is to compare the percentage of titles available in a given subject to the percentage of titles used and the percentage of usage. For example, while titles in the Social Sciences make up 30% of available content on ebrary, they account for 34% of titles with a session and 40% of sessions. Conversely, 30% of the titles available are in STEM disciplines, but these titles account for 29% of the titles with a session and only 26% of sessions overall. Arts and humanities titles represent 29% of the content available, 32% of the titles with a session, and 30% of sessions – much closer to what one would expect given availability. In most measures of use this pattern holds true, with social sciences titles used at a higher rate than would be expected and STEM titles used both at a lower rate and less intensely than would be expected.
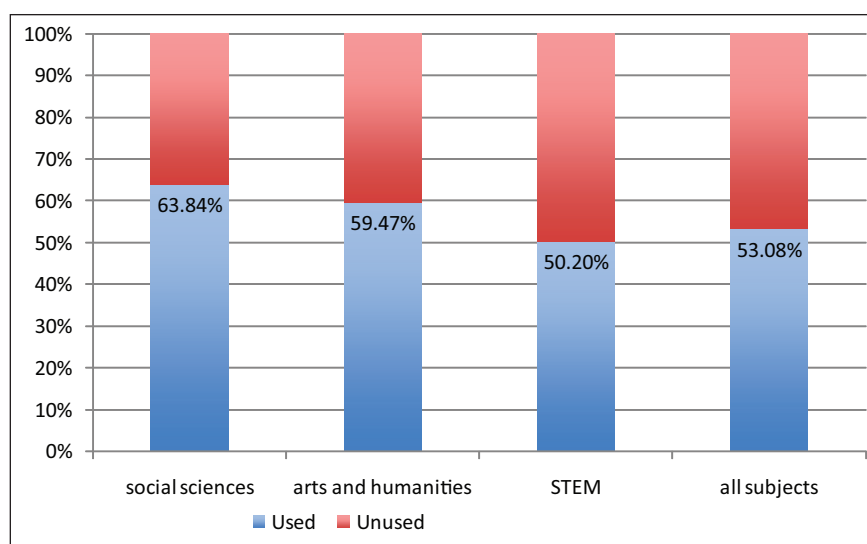
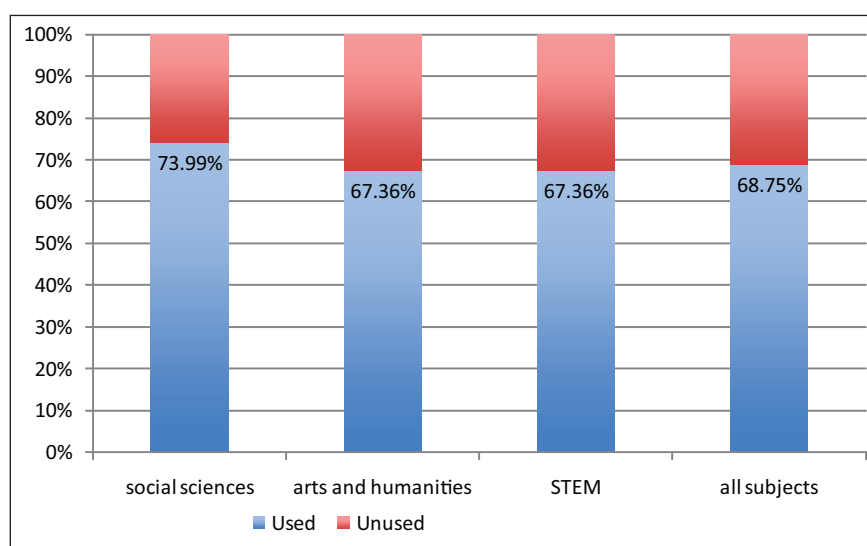Figure 7. Percentage of titles used, by discipline: ebrary



Figure 8. Percentage of titles used, by discipline: EBL

## Usage patterns per session

While broad measures of use can be valuable, in many ways what happens during any given interaction with the book is more telling. One user might examine a book quickly, glancing at one or two pages, while another might spend hours reading hundreds of pages. At the level of the individual user, those patterns are likely specific to an information need of the moment, but aggregated across hundreds of thousands of titles, those patterns tell us about practices by subject and region.

At the most basic level, every e-book has a session, and once a user begins a session she or he can read one or more pages of the book. The usage data show how many pages have been viewed in a particular book and also indicate how much time a user spent reading. In addition, the data include the number of downloads, the number of pages printed, and the number of pages with sections copied (and presumably pasted into a new document). Once a download occurs or pages are printed, the user is presumably reading offline, with no further measurement of use possible. Copying is more complicated, since it can be used to paste text into a new document for reading but can also be used for note taking or for pasting quotes into a paper. But all of these actions occur only after a session begins, so measuring these various actions per session allows for an understanding of how users typically interact with a book.

There are some interesting regional variations in terms of these actions per session. In the regions of the developing world – Africa, Asia Pacific,

'There are some interesting regional variations'

Latin America, Middle East and South Asia – there are both fewer pages viewed per session and less time spent in the book than in Australia/New Zealand, Europe, North America and the United Kingdom/Ireland. Conversely, there are more downloads per session in the developing world, with South Asia having almost twice as many downloads per session as the next closest region. It seems likely that downloading occurs in areas where internet connections are less secure, while reading on the aggregator platform occurs where those connections are more stable. Data from the International Telecommunications Union about internet access bears this out[3]. Printing and copying are less clearly differentiated, with the most printing occurring in the Middle East and the least in Latin America. Copying happens most often in the Middle East and commonly in Latin America as well, but least often in South Asia.

Just as there is regional variation in terms of actions per session, there are disciplinary differences as well. Across most regions, STEM titles show the most page views per session and arts and humanities titles have the least. Interestingly, the same pattern holds true for downloads per session. There is more variation in terms of printing and copying, but a very clear opposite trend for time per session, with the most time spent in arts and humanities titles and the least in STEM titles. Users of STEM books look at a lot of pages and frequently download as well, but overall they spend less time in the book. Arts and humanities users look at less of the book and download less frequently, but they spend more time in the book overall, suggesting a more immersive reading experience.

'there are disciplinary differences as well'

At the subject level, there are some interesting variations. Readers of books in the LC Class T (Technology) look at more pages per session than anyone else except for readers of F (History of the Americas). The other history LC Classes are also high on the list of page views per session. T is right at the top of downloads per session, while the History classes are at the very bottom. Meanwhile, readers spend more time per session in the History classes than any others, but less time in Technology books than all but two. Users of History books look at a lot of pages, they spend a lot of time in the book, and they rarely download. Users of Technology books look at a lot of pages and download a lot but spend very little time in the book.

## Usage of the most popular titles

In most libraries it is assumed that there is a long tail pattern to usage, with a relatively small percentage of titles accounting for a very high percentage of use. Examining just the top 10% of titles with a session allows for a test of this assumption for e-books. Worldwide, for ebrary, 64,314 titles accounted for about the top 10% of titles with a session, but accounted for 80.7% of all sessions. In all regions the top 10% of titles with a session accounted for 70% or more of sessions, and in all but two regions (Asia Pacific and the Middle East) those titles accounted for 80% or more of the sessions.

'a relatively small percentage of titles accounting for a very high percentage of use'

Worldwide, among the top 10% of titles, there were a disproportionately high number of titles in the social sciences, a higher than expected number in the arts and humanities and a lower than expected number in the STEM disciplines. Of the titles in the top 10%, 37% were in the social sciences, compared to 29% in the data set as a whole. For the arts and humanities, those numbers were 33% and 29% respectively, and for the STEM disciplines those numbers were 26% and 30%. Of the 15.8 million sessions that occurred in the top 10% of titles with a session, an even higher portion happened in the social sciences (42%), with fewer than expected occurring in the arts and humanities (27%) and STEM disciplines (27%).

There is a greater spread of usage outside the top 10% of titles in arts and humanities than in the other areas. For arts and humanities titles, 77.2% of all sessions occurred in this set of the top 10% of titles by usage. For STEM disciplines, 81.0% of all sessions occurred in the top 10%, and for social sciences, 83.1%. While the vast majority of sessions happen within this relatively small set of titles, the long tail for arts and humanities is a bit longer than for the other disciplines.

## Intensive vs. extensive use

At a very basic level, there are two ways that libraries measure usage: whether something gets used and how often that thing gets used. For a subject-level analysis, it is worth knowing how many e-books get used (extensive use) and how often those books are used (intensive use). In some subject areas, usage is both intensive and extensive (a high percentage of books get used at least once and typically those books get used often). In others, usage can be extensive but not intensive (lots of books are used, but typically not used often), intensive but not extensive (a high amount of use, but spread across a relatively small number of books), or neither.

Figures 9 and 10 show each LC Class (see Table 1 for reference) mapped onto a matrix comparing intensive and extensive use. The horizontal axis shows percentage of use (extensive use) while the vertical axis shows average amount of use (intensive use). The upper right quadrant includes subjects with both intensive and extensive use. In the lower right quadrant are those with extensive but not intensive use. The upper left quadrant depicts those with intensive but not extensive use. The lower left quadrant includes subjects with neither.
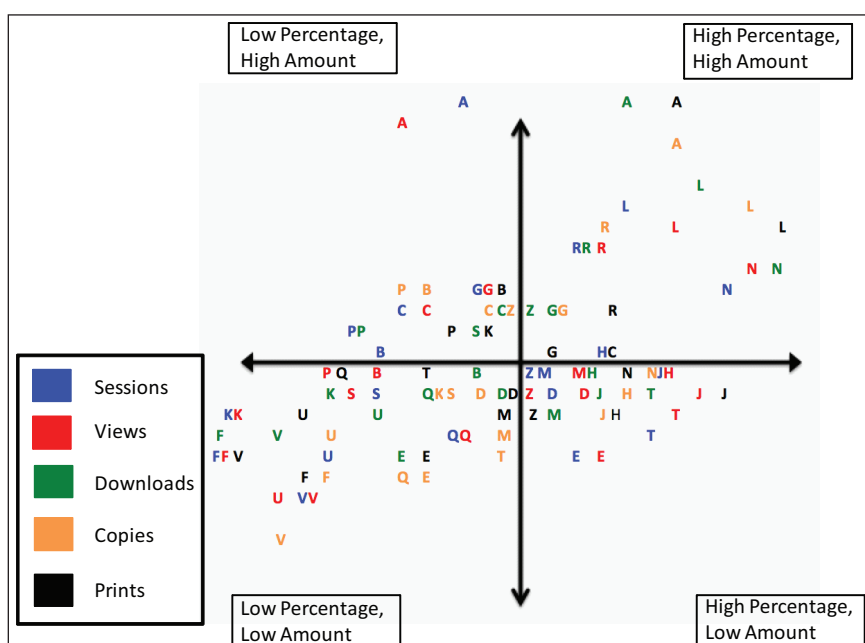


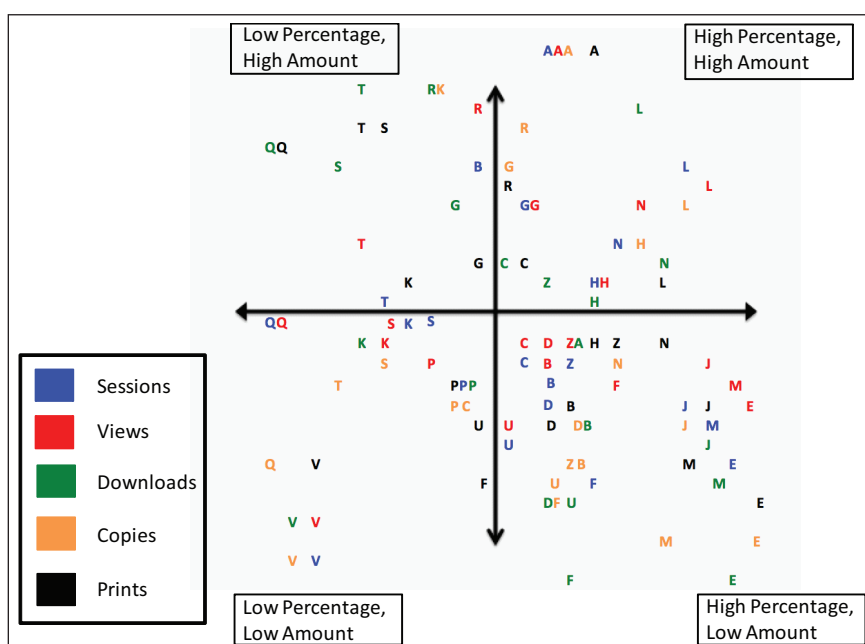Figure 9. Intensive vs extensive use: EBL



Figure 10. Intensive vs extensive use: ebrary

Though the degree of usage varies a bit between ebrary and EBL, the subjects tend to map onto the intensive-extensive grid similarly. L (Education) and N (Fine Arts) are the two subjects that are most consistently used both intensively and extensively. Books in both of these subjects tend to be in the middle of the pack in the measures of actions per session, suggesting that multiple measures of use are important for assessing the value of library resources.

## Conclusions

There are many ways to measure use, and this study explores just a few of them. The presentation at UKSG, available on Slideshare[4], and the published white paper[5], explore additional modes of analysis and provide more detail than is possible in a brief article. It is clear that there are variations by subject and by geographic region in how e-books are used. It is also clear that using multiple forms of analysis provides a more nuanced view of usage and value than just one. By some measures, books in the humanities perform poorly, for instance, while in others they perform quite strongly.

'using multiple forms of analysis provides a more nuanced view of usage and value'

The key finding in terms of geographic differences in usage is that users in the developing world are more likely to download e-books than to read them online, with the opposite pattern occurring in the developed world. Readers in Australia/New Zealand, Europe, North America, and United Kingdom/Ireland look at more pages and spend more time in the book online than those in the rest of the world. This is potentially linked to availability of stable internet connections and the need to download to read later in developing countries.

At the subject level, several interesting patterns are worth noting. Looking at broad disciplinary breakdowns (arts and humanities, social sciences, STEM), it is clear that social sciences titles are more likely to be used, and are used at a greater rate, than would be expected given their availability. STEM titles are used less than would be expected. These patterns hold true for the 10% of titles that are most heavily used. STEM titles, however, tend to do very well in terms of page views and downloads per session while arts and humanities titles do well in terms of time per session.

Zooming in a bit further, to the subject level, four subjects are worth highlighting. History titles account for many page views per session, very few downloads, and lots of time spent in the book, suggesting that readers of these books do immerse themselves in the book. Technology titles also account for many page views per session, but lots of downloads as well. Readers of these books, though, spend very little time in the book, suggesting that they are skimming, looking at lots of pages, and quickly getting the information they need. Education, a subject that is right in the middle of the pack in terms of all three of these measures, is the strongest performer by another measure: intensive and extensive use. Readers of Education books do not spend a lot of time in the book in any given session, nor do they look at a lot of pages, but they do look at a lot of books and they conduct many sessions overall. Finally, Fine Arts books have many pages viewed per session, but readers spend a relatively small amount of time doing so. As with Technology, readers look at many pages, perhaps examining images, but do not spend a lot of time doing so. Like Education, Fine Arts is an area with both intensive and extensive use. These four subject snapshots show that a single measure of use does not tell the complete story. Multiple modes of analysis show that there is value in all of these subjects, value that might be missed if a single metric were employed.

Usage patterns at the global level give a sense of general patterns and trends by subject and by region of the world. While it is valuable to know that users of History e-books are conducting immersive reading, it is more important to understand how that knowledge can inform local practice. It is hoped that this study will aid librarians in building e-book collections and serving their users.

Competing interests
The author has declared no competing interests.

References

1. Levine-Clark, M, *Global Trends in eBook Usage: Patterns from 10,000 Libraries*:
   http://go.proquest.com/globalbookusage2015 (accessed 5 June 2015).

2. Levine-Clark, M and Paulson, K, 'E-Book Usage on a Global Scale: Patterns, Trends, and Opportunities', UKSG Annual Conference, Glasgow, 30 March–1 April 2015:
   http://www.slideshare.net/MichaelLevineClark/levineclark-michael-and-kari-paulson-ebook-usage-on-a-global-scale-patterns-trends-and-opportunities-uksg-annual-conference-glasgow-march-30april-1-2015 (accessed 4 May 2015).

3. International Telecommunications Union, *Key ICT indicators for developed and developing countries and the world (totals and penetration rates)*:
   http://www.itu.int/en/ITU-D/Statistics/Documents/statistics/2014/ITU_Key_2005-2014_ICT_data.xls (accessed 4 May 2015).

4. Levine-Clark, M and Paulson, K, ref. 2.

5. Levine-Clark, M, ref. 2.

Michael Levine-Clark

Associate Dean for Scholarly Communication and Collections Services
University of Denver, 2150 East Evans Avenue, Denver, CO 80208, USA
E-mail: michael.levine-clark@du.edu

ORCID ID: http://orcid.org/0000-0003-3001-0261