

## Does Format Matter? Reader Preferences in an Academic Library Context

Jennifer L. Robertson  
*University of Toronto Libraries*, jen.robertson@utoronto.ca

Weijing Yuan  
*University of Toronto Libraries*, weijing.yuan@utoronto.ca

Marlene van Ballegooie  
*University of Toronto Libraries*, m.vanballegooie@utoronto.ca

Follow this and additional works at: <https://docs.lib.purdue.edu/charleston>



Part of the [Collection Development and Management Commons](#)

An indexed, print copy of the Proceedings is also available for purchase at:

<http://www.thepress.purdue.edu/series/charleston>.

You may also be interested in the new series, Charleston Insights in Library, Archival, and Information Sciences. Find out more at: <http://www.thepress.purdue.edu/series/charleston-insights-library-archival-and-information-sciences>.

---

Jennifer L. Robertson, Weijing Yuan, and Marlene van Ballegooie, "Does Format Matter? Reader Preferences in an Academic Library Context" (2015). *Proceedings of the Charleston Library Conference*. <http://dx.doi.org/10.5703/1288284316263>

# Does Format Matter? Reader Preferences in an Academic Library Context

*Jennifer L. Robertson, EResource & Mathematics Web and Scholarly Communications Librarian, University of Toronto Libraries*

*Weijing Yuan, Head, Licensing & eResource Acquisitions, University of Toronto Libraries*

*Marlene van Ballegooie, Metadata Librarian, University of Toronto Libraries*

## Abstract

Although many academic libraries have dramatically increased their e-book acquisitions in recent years, questions linger about format preference. When a scholarly monograph is made available in both print and electronic formats, which format will users prefer? Does format even matter? At the University of Toronto Libraries, we analyzed usage data for scholarly monographs from three key university presses, covering thousands of titles over several years of publication. By comparing print and e-book usage patterns of identical titles, our goal was to examine format preferences and determine if there are differences in usage across subject disciplines or publishers. Through this analysis, our aim is to question whether continued acquisition of the same content in multiple formats is necessary and desirable, especially in an era of rapid technological change, increased pressure on library acquisitions budgets, and diminishing physical storage space.

## Introduction

Integrating e-books into a library's collection development strategy requires understanding of how they are used relative to their print counterparts. When a scholarly monograph is made available in both print and electronic formats, which format will users prefer? Does format even matter? Are the same titles used in both formats? Are e-books used more as they are more accessible? Do usage patterns vary by subject or publisher? This study tried to answer these questions by analyzing usage data for e-books from three key university presses, covering thousands of titles over several years of publication, which were available in the University of Toronto Libraries collections.

The University of Toronto Libraries is comprised of forty-four libraries on three campuses and has rich and diverse collections that serve the teaching and research needs of the University's 67,000 students and over 7,000 faculty members. The collection development department in the central library system usually licenses electronic resources for the entire University. Other libraries have acquisitions budgets independent of the central library to meet local needs.

Over the last decade and a half, the library's e-book collections have grown tremendously and there are now close to 1.7 million e-books. The library has focused on licensing e-books for perpetual access, with unlimited concurrent users and without restrictive DRM whenever possible, and purchasing frontlist packages directly from publishers. For content in humanities and social sciences, even where frontlist e-book collections were licensed, print copies were still provided through the central library's approval plans as long as they fit the profiles. As a result, the library has built a large pool of titles with print and electronic equivalents.

## Scope and Methodology

This book study focused on analyzing the usage of 6,555 print and online equivalents for three university presses: Oxford University Press, Cambridge University Press, and Duke University Press. For Oxford University Press, the study was limited to 3,624 Oxford Scholarship Online (OSO) titles with imprint years between 2010 and 2014. In the Cambridge Books Online collection, we analyzed 2,121 frontlist titles with imprint years between 2012 and 2014. Finally, for Duke University Press, 810 titles from the e-Duke Book

Scholarly Collections with imprint years between 2008 and 2014 were considered for analysis. In all three collections, the backlist titles were excluded from the study due to the fact that the print and electronic versions were not available to users simultaneously.

Print data collection centered upon batch searching ISBNs against our Sirsi ILS system using a Z39.50 client. Once the resulting MARC records were de-duplicated and verified for correctness, the basic bibliographic elements were extracted from the MARC records. To obtain data pertaining to the individual copies of the print titles, custom API queries were used to retrieve the call numbers, number of physical copies, the circulation counts, and the date of last circulation. In cases where the library had multiple copies of a particular title, the data was aggregated to represent the total usage for the title. For the electronic data collection, we downloaded the COUNTER BR2 reports for all available years and merged them into a master spreadsheet for each vendor. This allowed for the calculation of monthly and yearly totals. Once all of the electronic usage data was collected, the ISBN was used as the match point to combine the print and electronic usage data into one master list.

## Limitations

In the data collection process we encountered some limitations worth noting. The first limitation was that our print and electronic use data was not fully compatible. With the print data, we were only able to obtain the total charges for a title and the last date that the book was charged. Therefore, while we had a record of electronic usage by month and year, we could not get comparable statistics for the print equivalent. Inconsistencies in the COUNTER reports were also noted in the data collection process. For example, through the course of the study both Duke and Oxford switched from COUNTER 3 to COUNTER 4, and Duke migrated platforms from ebrary to Highwire, which likely affected how different vendor sites counted a "use." For this study, we consider Oxford, Cambridge, and Duke on the new Highwire platform to be counting chapters or equivalent, and Duke on ebrary to be counting all

uses as a section. Another limitation of the COUNTER BR2 reports is that only used titles were included in the report. Without a full list of titles available on each platform, we had to make the assumption that the electronic version was available at same time as the print version and that if a title was missing from the BR2 reports this indicated that there was zero usage. In the electronic data collection process we also observed that COUNTER data was not available for the full date range of each press. For example, although we had print and online equivalents for Oxford since 2010, we could not get COUNTER data until March 2011. A final limitation worth mentioning is that the content we have locally loaded is unaccounted for in the study. For each of the university presses considered, we have an equivalent set of e-books loaded on local servers. However, due to the fact that we could not obtain compatible statistics for this content, we were required to exclude the locally loaded versions of the e-books from the study.

## Results

### *Title Use by Format*

Overall, close to 80% of books were used at least once, and patrons at the University of Toronto showed a slight preference for print. Print books were used 1.16 times more frequently than online books, where 66% of print books and 57% of online books were used at least once. Looking at the presses individually, Oxford print books were used 1.45 times more frequently than online books, showing a much stronger tendency to choose print. Use of Cambridge books was more balanced with a slight preference for online books, which were used 1.16 times more frequently than print, and Duke's print books were chosen only slightly more often than online. Oxford had the highest percentage of unused books at 24%, closely followed by Cambridge at 22%, while Duke's collection was very highly used with almost 90% of Duke's titles used at least once. Furthermore, Duke had the highest percentage of books used in both formats at close to 70%, with Oxford and Cambridge trailing behind at between 40–45%.

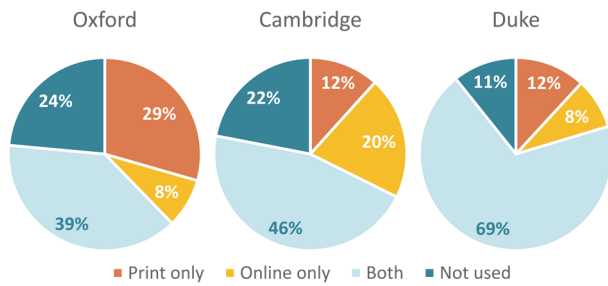


Figure 1. Title use by format.

### Number of Titles by Publication Year

These column charts display the number of titles used by format and publication year. Looking at each press, the number of used print and online books progressively decreased in the last three publication years, most likely because newer books have had less time to be discovered and used. For Oxford, the gap between used print and online titles widened unexpectedly for the 2013 and 2014 imprint years, and Duke showed a larger than expected decrease in used online titles for 2013.

### Percentage of Titles by Year Last Charged

For print books that circulated at least once, 77% were checked out in the last year, and the last two years accounted for over 90% of all title circulations, ranging from 89% for Oxford to 97% for Cambridge. So print books are still circulating, online has not replaced print, and most books that haven't been used at all continue to be used.

### Total Use by Publication Year

Between the three presses, there were 30,805 print circulations and 351,752 online section views or downloads during the study period. Oxford had the highest number of circulations, but also the largest collection of books. Duke had the highest online use despite its smaller collection size, but this is likely due to the earlier start date and the difference in how the ebrary platform counts use. Use for Duke was also significantly smaller for the 2013 and 2014 imprint years and, while we expect newer books to have generated less use, this very large decrease is also a likely side effect of the platform migration.

### Percentage of Print and Online Use by Range

Print and online use was also grouped into predefined ranges. For print books, the majority of use fell between 2 to 5 circulations with a very small proportion of use at 21 and over. For online books, the majority of use fell between 1 to 10 uses with a very small proportion of use generated at 51 and up. In both formats, Duke had a much larger proportion of books in the highest ranges.

### Format Preference by Subject

After examining each university press collection as a whole, we analyzed them by subject.

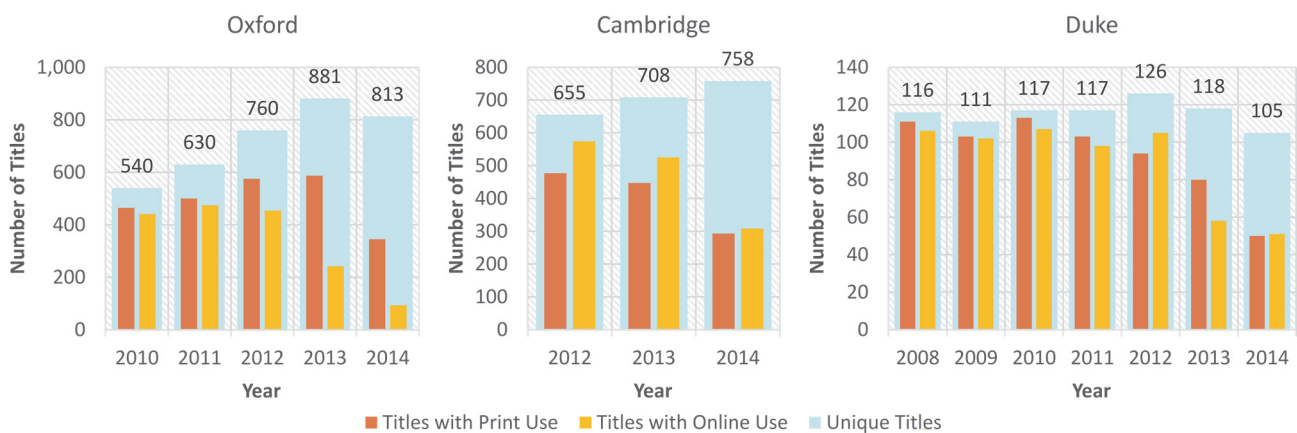
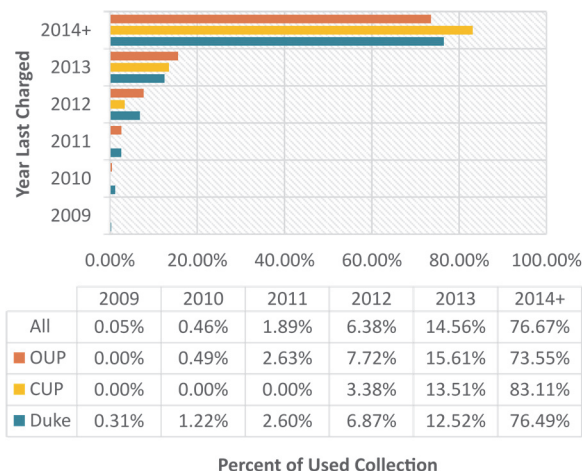


Figure 2. Number of titles by publication year.



**Figure 3. Percentage of titles by year last charged.**

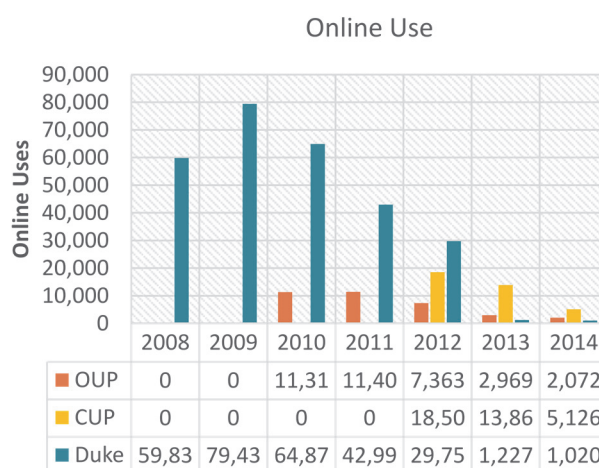
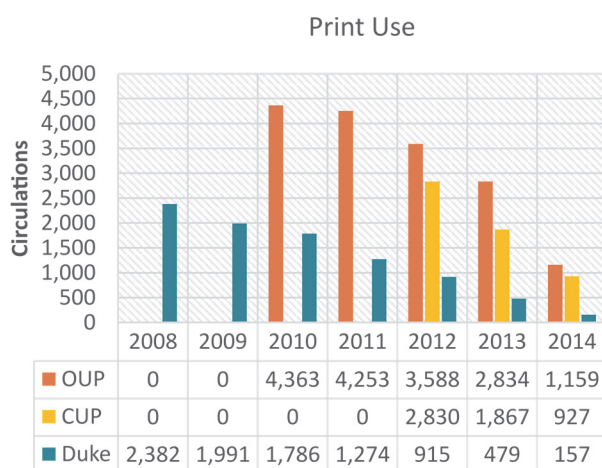
Categorizing titles by LC class produced 20 collection subsets for Oxford and Cambridge and 19 collection subsets for Duke. For all collections, subjects with less than 11 titles per subject were excluded.

To analyze the amount of use for both formats by subject, we focused on relative use. Previously we compared counts of used and not used books in both formats. With relative use, we compared the proportion of use for a subject relative to its size. A relative usage factor was calculated separately for each subject and format by dividing the proportion of use by the proportion of the subject within the collection.

Next we compared each relative usage factor to its expected use to determine whether a format

generated use above or below the expected value. For example, if a subject's size takes up 25% of the total collection, it is expected to generate 25% of the collection's total use, and when this happens the relative usage factor equals 1. Where factors were greater than 1, the subject was well used in that format and generated more use than expected. For factors less than 1, the subject was underused. The amount a relative usage factor diverges from 1 indicates the percentage at which the subject is over- or underperforming. For example, for a subject with a relative usage factor of 1.20, use was 20% higher than expected, or 120% of its expected use, and for a subject with a relative usage factor of 0.70, use was 30% lower than expected, or 70% of its expected use.

After calculating the relative usage factors for each subject and format, we plotted the factors for online use against the factors for print use by subject on a scatterplot to view the overall pattern. Horizontal and vertical lines were inserted at 1 to indicate the expected use for each format. Comparing relative usage factors across all three presses, we found that for most subjects the use of one format was tied to the use of the other. Many subjects clustered around the expected use; if a subject was popular, use tended to be high for both formats and if unpopular, low for both formats. There were, however, subjects for each press that had relatively high or low use in one format only. For example, Oxford Technology and Cambridge Medicine had much



**Figure 4. Total use by publication year.**

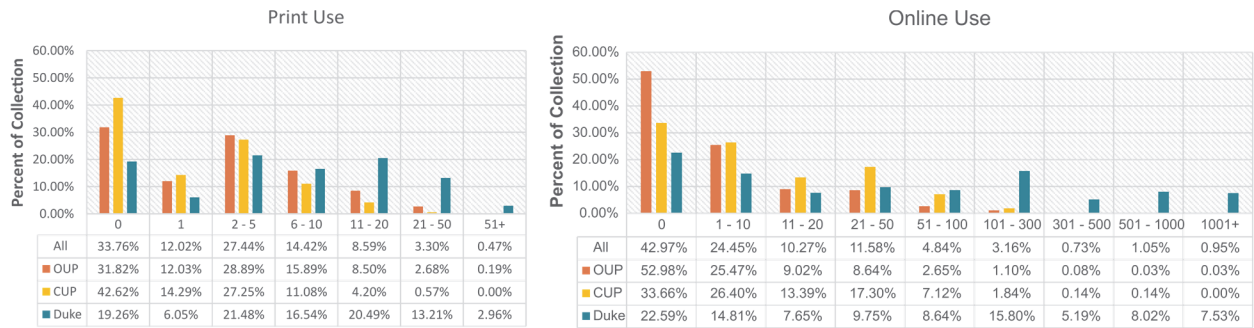


Figure 5. Percentage of print and online use by range.

higher use than expected for online and less than expected use for print.

### Difference Between Print and Online Usage Factors

To explore format preference in more detail and help identify subjects where user preference for one format over the other was strong, we analyzed the gap between print and online relative usage factors by plotting the difference on a bar chart. Each bar represents a subject and the bar color indicates the preferred format. The endpoints of each bar are equal to the factors themselves, so the right edge of each bar is the factor for the preferred format and vice versa. The length of each bar is equal to the gap between the subject's print and online relative usage factors. Therefore, the longer the bar the more preference was shown for one format over the other.

Several subjects stood out either because of the substantial gap between print and online relative usage factors or the large distance from the line

defining expected use. Some subjects seemed popular with the entire bar well to the right of the line of expected use, while others had bars well to the left and were underused in both formats. For Oxford, print exceeded online in 11 of 17 subjects, while the split was more even for Cambridge and Duke with just over half of subjects with a print preference. Where there were moderate to large gaps in relative format use, it was not the same subjects across all three presses, and a format preference exhibited by one press could be the opposite for another press. And while Oxford and Cambridge had several subjects with large gaps in format preference, Duke's relative usage factors for print and online tended to correspond more closely with gaps not nearly as large, which is consistent with Duke's more even format use.

## Discussion and Conclusions

### Does Format Matter?

When a book is available in both print and electronic formats, which format will users

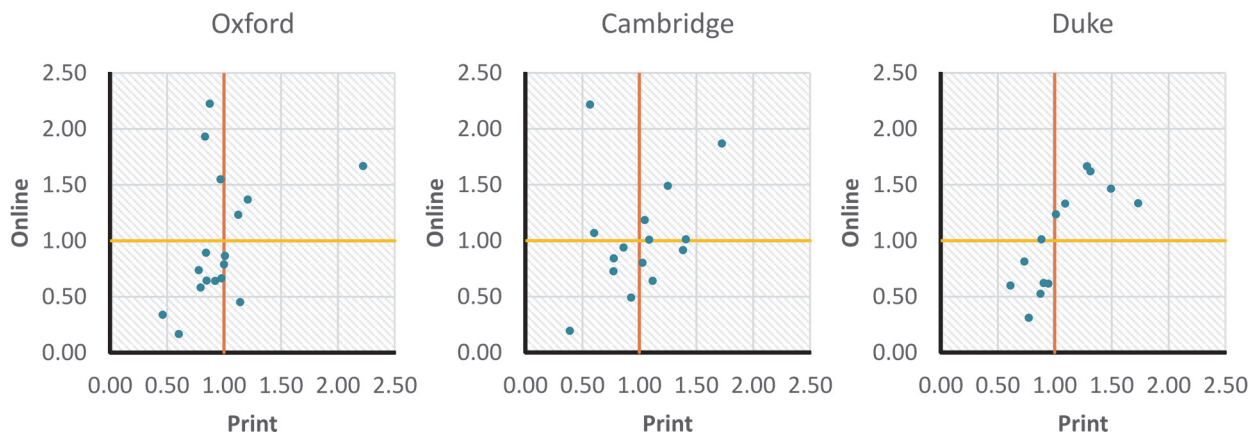
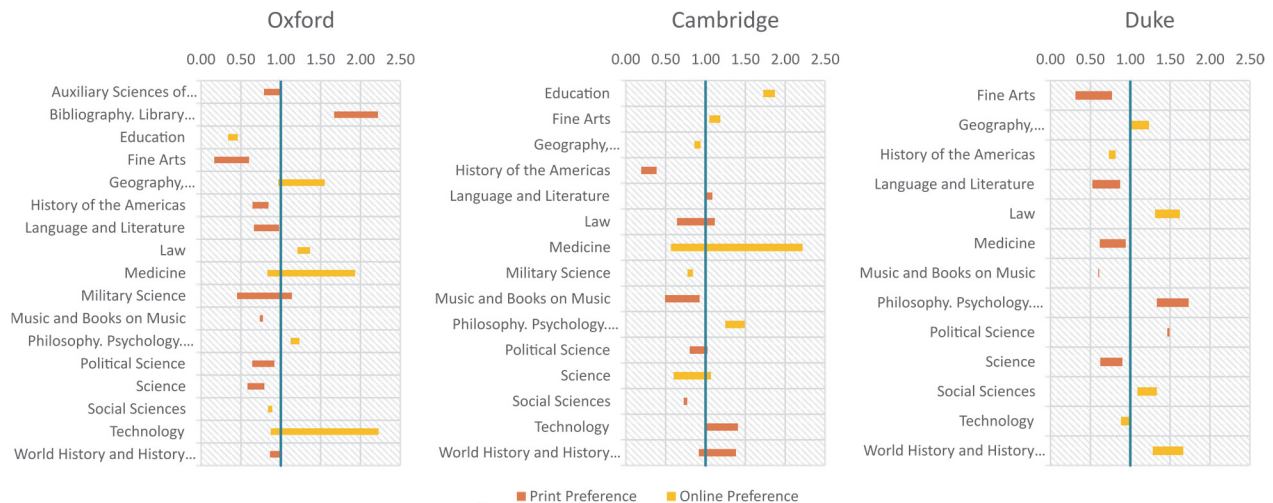


Figure 6. Format preference by subject: print versus online usage factors.



**Figure 7. Format preference by subject: difference between print and online usage factors.**

prefer? Although we came into the study with the assumption that our users would prefer electronic, by the end of it we were not entirely convinced. Our study shows that both formats were used quite extensively. For all of the presses considered, most titles had use in one format or another. Within the used portion of the collections, between 39% and 69% of titles were used in both formats. The circulation data confirms that print titles continue to be used several years after being released. In fact, the majority of print books circulated within the past 2 years. These statistics indicate that the presence of an electronic version of a book does not automatically displace the print title on the shelf.

#### *Publisher Data, Knowledge Bases, and E-Book Use*

In the data analysis, we observed that a substantial number of titles were used in one format only. For example, Oxford had an abnormally high degree of “print only” usage, and electronic usage in the last two years was comparatively lower than the other presses. A probable explanation for the low online usage of OSO content is that a large number of titles from 2013 and 2014 were missing from the Oxford Scholarship Online Frontfile package that we had activated within our electronic resource management system. In the University of Toronto’s e-resource management context, the majority of catalog records for e-resources

emanate from a MARC record service that is connected to titles activated within the knowledge base. As a result of the data omission within the knowledge base, we see a chain reaction that ultimately leads to no records in our library catalog and contributes to low use of the resources. Uncovering an anomaly such as this highlights the importance of a well-functioning supply chain for e-resource metadata and illustrates how critical it is in promoting the use of electronic resources.

#### *Measuring Use*

Throughout both the data collection and analysis processes we were continually faced with the question: Can the statistics truly gauge user preferences? When analyzing print and electronic usage, the statistics are not totally comparable. With print we are focusing on circulations of entire books which generally cover a long range of time. In contrast, electronic resource accesses tend to cover shorter periods of time and focus on chapters or sections of books, not books in their entirety. Statistics also cannot measure the intensity of use. Whether an item is used extensively or whether it is clicked on once and never consulted again, we really cannot gain any insight into this from the statistics. Similarly, by focusing print use on circulations, we only capture use that happens outside of the library walls when the patron physically signs out the book; we do not account at all for the in-house use of print items. The uniqueness of a user is also

unaccounted for in use studies. Due to privacy restrictions, we cannot tell from the statistics whether many users are accessing an electronic book or one user repeatedly. Concerns like these are not new; they are well documented by previous studies measuring print and online use.

Although the statistics serve as an indication of use, we should not read into the statistics too deeply. The value of a study such as the one we conducted is not necessarily in the absolute numbers. What is of greater significance is the overarching trends that are represented within the statistics. In our analysis, we found that determining the “relative use” was more useful than focusing on the actual numbers, as it told us more about how resources were performing relative to other subjects within the collection, as well as which format tended to prevail in particular subjects.

### *Future Research*

This study is one of the early steps in evaluating e-books at the University of Toronto Libraries. Through time, we hope to refine our process and use this study as a benchmark for future investigations. Due to the limitations of a purely statistical approach to use analysis, for future iterations of this study we are considering the inclusion of qualitative data to paint a better picture of user preferences. We hope that through surveys and focus groups we can gain more insight into how both print and online formats are used by patrons. Data in this study clearly showed that our patrons used both formats and in many cases preferred one format over another. It is therefore reasonable to conclude that format does matter. However, understanding why print books were still heavily used while e-books were also an option is a more complex question that warrants further investigation.