

PREPRINT

Published in: Issues in Science & Technology Librarianship, Winter 2018.

<http://www.istl.org/18-winter/refereed4.html>

DOI:10.5062/F4DV1H4T

## **Analysis of Citations to Books in Chemistry PhD Dissertations in an Era of Transition**

David Flaxbart  
Chemistry Librarian  
University of Texas at Austin  
[flaxbart@austin.utexas.edu](mailto:flaxbart@austin.utexas.edu)

### **Abstract**

A citation analysis of chemistry PhD dissertations at the University of Texas at Austin yielded data on how often graduate students cite books in their bibliographies, and on the characteristics of the books cited, in terms of age and local ownership. The analysis examined samples of dissertations selected from five discrete years – 1988, 2006, 2009, 2012, and 2015 – in order to provide longitudinal data on how citation trends are changing during a transition period in libraries. Data indicated that chemistry graduate students cited low numbers of books relative to journal articles, confirming expectations from similar studies, and that the trend over the time period studied is downward, despite the increasing number and availability of e-books. The results could inform collection management decisions and strategies for promoting book content to graduate students in future.

### **Introduction**

Libraries at PhD-granting institutions have spent significant resources on acquiring, storing, and maintaining local collections of monographic materials (commonly referred to as “books”), traditionally in print format but also more recently in digital formats. In the STEM fields, where peer-reviewed journal literature is regarded as the primary measure of academic success, libraries have rightly focused on maximizing access to journal content while treating other forms of literature as secondary. At the same time,

this bias has been reinforced, intentionally or not, by libraries' trending away from housing large print collections in valuable real estate in close proximity to their patrons. Academic libraries are deep into a profound transition, from the traditional role of acquiring and storing physical collections in anticipation of need, to providing optimal (just-in-case or just-in-time) access to a vast universe of external digital content. What is not clear is whether these changes *reflect* changing user needs, or are actually driving those changes.

The aggregate usage of a library's book collection by graduate students can be analyzed through patron-type circulation statistics over time, but this approach is fairly one-dimensional. Citation analysis of PhD dissertations offers a more granular approach to studying their actual use of literature. In STEM fields, most citation analysis has focused on use of the journal literature, which accounts for the bulk of citations and financial investment. This study turns the spotlight to "books" – broadly defined here as non-serial, non-peer-reviewed publications by one or more authors or editors, and taken to encompass single-author monographs, edited multi-author collections of chapters, and textbooks at all levels, in both print and electronic formats; but to exclude explicit conference proceedings, technical reports, theses and dissertations, and other one-time works. While books account for only a small portion of a typical research library's spending, print books have disproportionately high maintenance costs due to higher circulation and demand compared to less-used archival runs of printed journals.

The discipline under scrutiny is chemistry, the so-called "central science," well known as a highly literature-dependent field where the supremacy of traditional peer-reviewed journals is as yet unchallenged. The study sought to answer these questions: How often do chemistry graduate students cite books in their dissertations? Is the rate changing over time? What is the age profile of books cited? What proportion were available in the library's collection? How do the results compare to similar analyses at other institutions? Answers could help libraries plan for future acquisition strategies, and inform the management and accessibility of existing collections.

Librarians have used citation analysis of scientific publications for decades to inform collection development decisions (Subramanyam 1980; Kushkowsky et al. 2003). The application of citation analysis methods to locally produced PhD dissertations is also well represented in the literature, typically focusing on specific fields of study (Edwards 1999; Eckel 2009). Most analyses have tended to dwell on citations to journals, which are the most commonly cited form of literature in STEM fields, and understandably the locus of most collection development and funding attention in academic science and technology libraries. In collection management situations, the format and age of materials cited by local authors have also received attention (Ortega 2008). For chemistry in particular, several citation analyses of dissertations are relevant to the present one. While they too focused primarily on citations to journals, they included summary data on the citations to monographic materials as well (Mubeen 1996; Gooden 2001; Vallmitjana & Sabaté 2008; Kayongo & Helm 2012; Zhang 2013; Gohain & Saikia 2014). None of these studies, however, used a longitudinal approach to look at changes over time.

Most citation studies, including this one, focus on dissertations authored at a single institution. While this maximizes relevance to that institution's own situation and decision-making, it carries the limitation of not necessarily being applicable across multiple institutions. The method, while somewhat tedious to carry out at sufficient scale, is simple enough to be replicated readily at other institutions interested in analyzing their own students' research behaviors.

The University of Texas at Austin (UT Austin) is a large, R1 university with over 11,000 graduate students (out of 51,000 total students) and prominent graduate programs across all STEM fields. Its highly ranked chemistry and biochemistry programs award an average of 36 PhD degrees per year. [Note 1] Graduate students have been required to submit electronic theses and dissertations (ETDs) to the [Graduate School](#) since 2004, and with few exceptions hardcopy dissertations are no longer added to the library's collection. Students upload their dissertations into the [Vireo](#) online system hosted by the Texas Digital Library (TDL) on behalf of the Graduate School. Most chemistry students choose to invoke an optional embargo period that blocks their ETDs from public access

for at least two years after completion. [Note 2] After final approval and expiration of any embargo, the ETDs are cataloged and deposited in the institutional repository, known as [Texas ScholarWorks](#), and are openly available thereafter.

Doctoral theses in STEM fields are not always ideal vessels of scholarship. They have been the target of angst, complaints, and calls for change for many years, yet they remain largely unchanged (Anonymous 2016). Chemists in particular often regard the dissertation as a mere degree compliance requirement, rather than a valuable scholarly work in its own right. Most of the useful research described in a dissertation has already been, or soon will be, published in peer-reviewed journals. Furthermore, as a single-author tome, the dissertation does not accurately reflect the interactive, team-oriented nature of modern scientific practice (Gould 2016). Nevertheless, the student is expected to demonstrate mastery of the subject and its associated literature, so the dissertation's bibliography tends to be exhaustive and provides a thorough record of the student's background research, and therefore is a useful snapshot of one individual's use of the information content provided by the library.

## **Methods**

Using lists of bibliographic records generated from UT Austin's local library system (Innovative's Sierra), the author used an online random number generator to select samples of ten Chemistry and Biochemistry ETDs from each of four completion years, at three-year intervals: 2006, 2009, 2012, and 2015. These years and interval were chosen to provide longitudinal data across the time period after the mandatory adoption of electronic submission. In addition to the ETDs, ten hardcopy dissertations from 1988 were randomly selected to provide a pre-ETD, pre-Web control group for purposes of baseline comparison. Table 1 shows the size of the sample relative to the total number of dissertations identified for each year. (The lower total for 2015 is most likely due to the two-year embargo that many chemistry students opt to place on their dissertations. This results in long delays in cataloging and ingest into the local repository.)

Table 1. Dissertations identified by sample year.

	1988*	2006	2009	2012	2015
Total Identified	36	28	35	41	15
Sample %	27.8%	35.7%	28.6%	24.4%	66.7%

\* 1988: Print dissertations; 2006-2015: Electronic dissertations

The author downloaded each selected ETD document from the university's repository and coded it by sub-discipline based on its title and the faculty adviser's departmental division. The categories assigned were Analytical, Biochemistry, Electrochemistry, Inorganic, Organic, and Physical/Theoretical (see Table 3). Finally, the bibliography section of each ETD was saved as a separate PDF file. The hardcopy 1988 dissertations were retrieved from storage and analyzed by hand.

The analysis of the bibliographies involved hand-counting unique bibliographic references to arrive at a grand total, while simultaneously counting and coding each unique reference according to a rubric of bibliographic formats:

- Journal article
- Book
- Monographic serial/review/annual (i.e., non-journal review-type publications that are typically cited as serials, such as *Methods in Enzymology*; *Annual Review of Physical Chemistry*; *Advances in Chemical Physics*, etc.)
- Conference paper, preprint or abstract (including *ACS Symposium Series*)
- Patent
- Dissertation/Thesis
- Unpublished
- Web site
- Personal communication
- Other (e.g. software; manuals; catalogs; unidentified, etc.)

Duplicate references, where evident, were counted only once.

Each reference in a bibliography identified as a “book” was coded for date of publication to enable calculation of the book’s age at the time of citing. Each book was also searched in the local catalog system to determine if the library held a copy in print, and if that copy was likely present in the collection at the time of the dissertation’s completion. If the latter was true, even if the book was now no longer available, the book entry was coded as “owned.” The presence of an e-book version of a title in the current catalog was ignored: due to constantly shifting e-book availability via various purchasing and demand-driven programs over the years, it was not possible to determine an e-copy’s history or availability at the time of the dissertation’s completion.

## **Results and Discussion**

The 50 dissertations analyzed yielded a total of 8,723 references, of which 474 (5.4%) were categorized as “books” (Table 2). The median age of the books as of the year of dissertation completion was 11 years. 436 of the 474 books (92%) were judged to have been available in the university’s library collection at the time of completion.

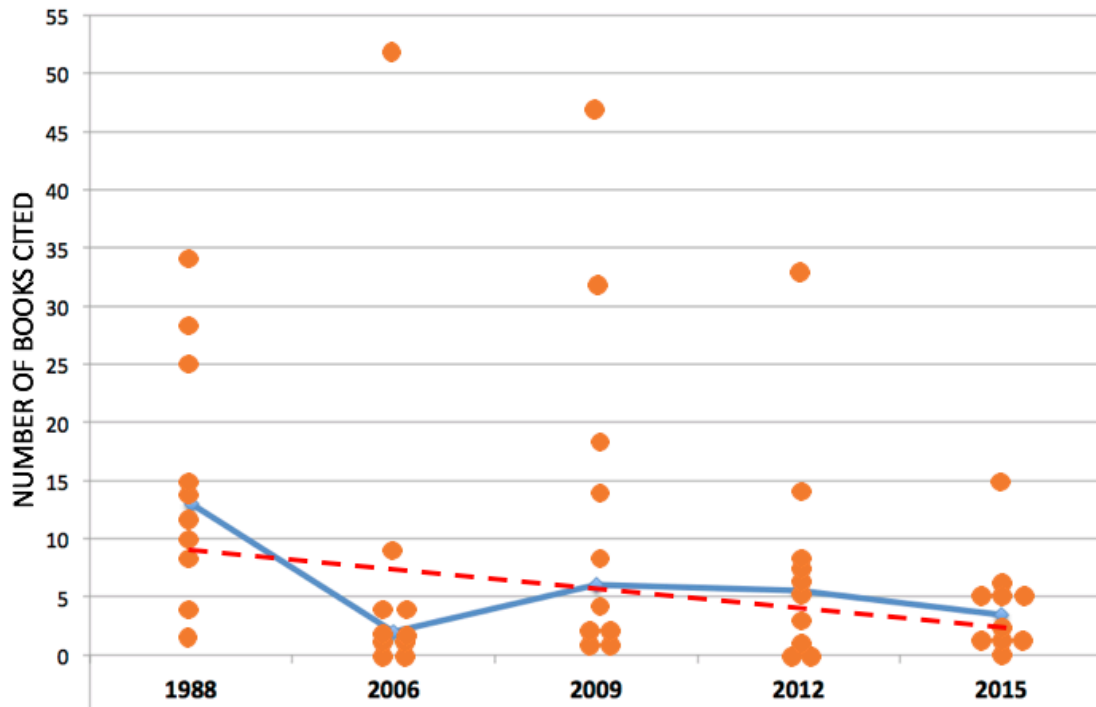
As chemistry is well-known as a journal-centric discipline, it was no surprise that over 90% of the references were to journal articles, with books coming in a very distant second, and all other categories combined totaling only 3 percent.

Table 2. Type of Work Cited, All Years

<b>Citation Type</b>	<b>Total</b>	<b>Percent</b>
Journal Article	7,980	91.5%
Book	474	5.4%
Monographic Serial	123	1.4%
Dissertation/Thesis	35	0.4%
Web Site	18	0.2%
Conference	15	0.2%
Patent	14	0.2%
Unpublished	10	0.1%
Personal Communication	3	0.0%
Other/Unknown	51	0.6%
Total	8,723	100%

Figure 1 plots the number of books cited (Y-axis) in each sampled dissertation (represented by a dot) by year of dissertation (X-axis). With 1988 as the pre-ETD baseline year, both the median number of books cited and the trend line (dashed line) are downward: newer dissertations tended to cite fewer books. With the exception of 1988, most students cited fewer than ten books in their dissertations, a trend clearly visible in this plot.

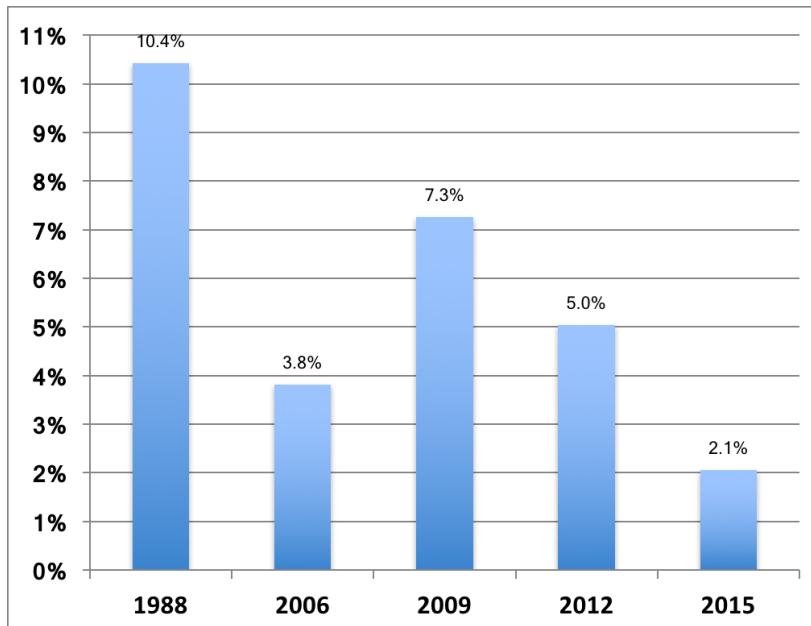
Figure 1. Number of Books Cited by Year of Dissertation



Even adjusting for the variable number of total references across the sample dissertations, books represent a declining proportion of the total over time (Figure 2). (The low figure for 2006 might be an artifact of sampling error.)

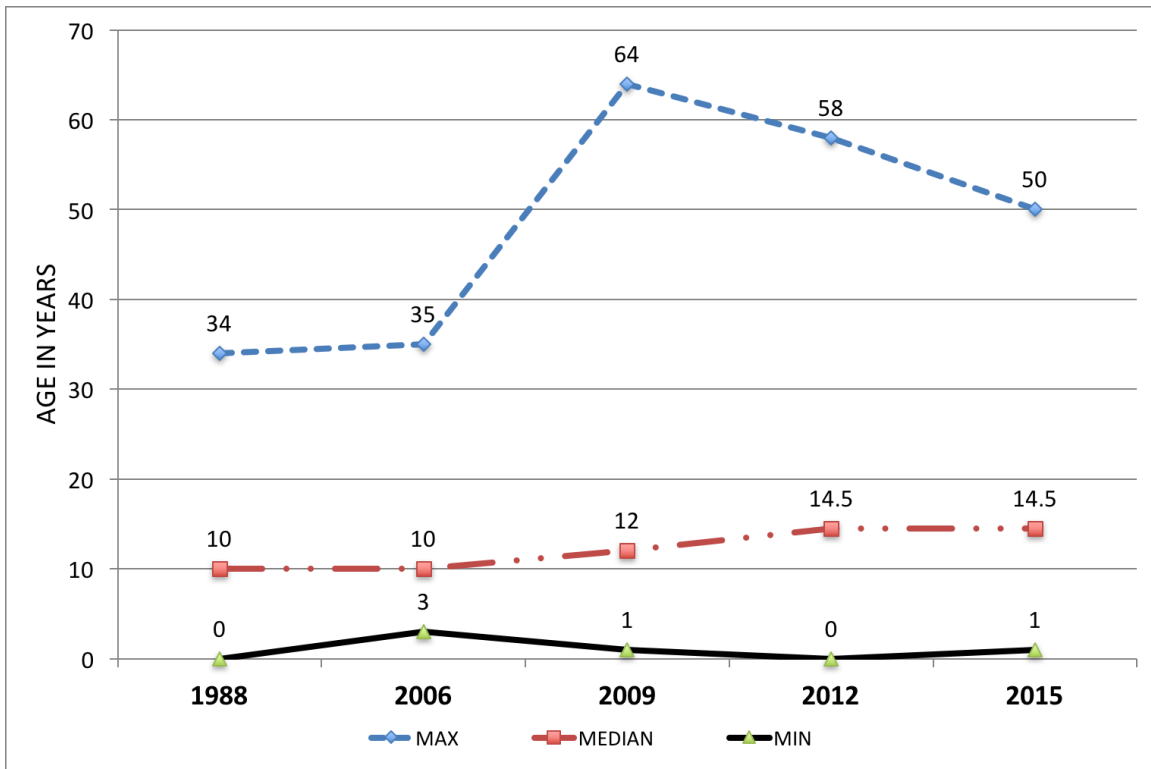


Figure 2. Books as Percent of Total References



The age of each cited book was calculated by subtracting the year of publication from the year of dissertation completion. The minimum age of books cited ranged from zero to three years. The median age of books cited increased from ten years (1988 and 2006) to 14.5 years (in 2012 and 2015), indicating that more recent students were citing somewhat older books. The maximum age of books cited varied more widely, ranging from 34 years in 1988 to 64 years in 2009 (Figure 3).

Figure 3. Age in Years of Books Cited, by Sample Year



It is a common observation that scientists cite recent literature more frequently than older literature, and this was borne out by the book age data shown in Figure 4. Across all the sampled years, books less than ten years old predominated, with the mode (i.e., the most frequently occurring value) being six years. Figure 5 depicts the cumulative percentage of books cited by age: the cited “half life” (50% of books cited, i.e. the median, represented by the red line) was 11 years, and 80% (Pareto distribution, represented by the black line) of all books cited were 21 years old or less.

Figure 4. Age Distribution of All Cited Books

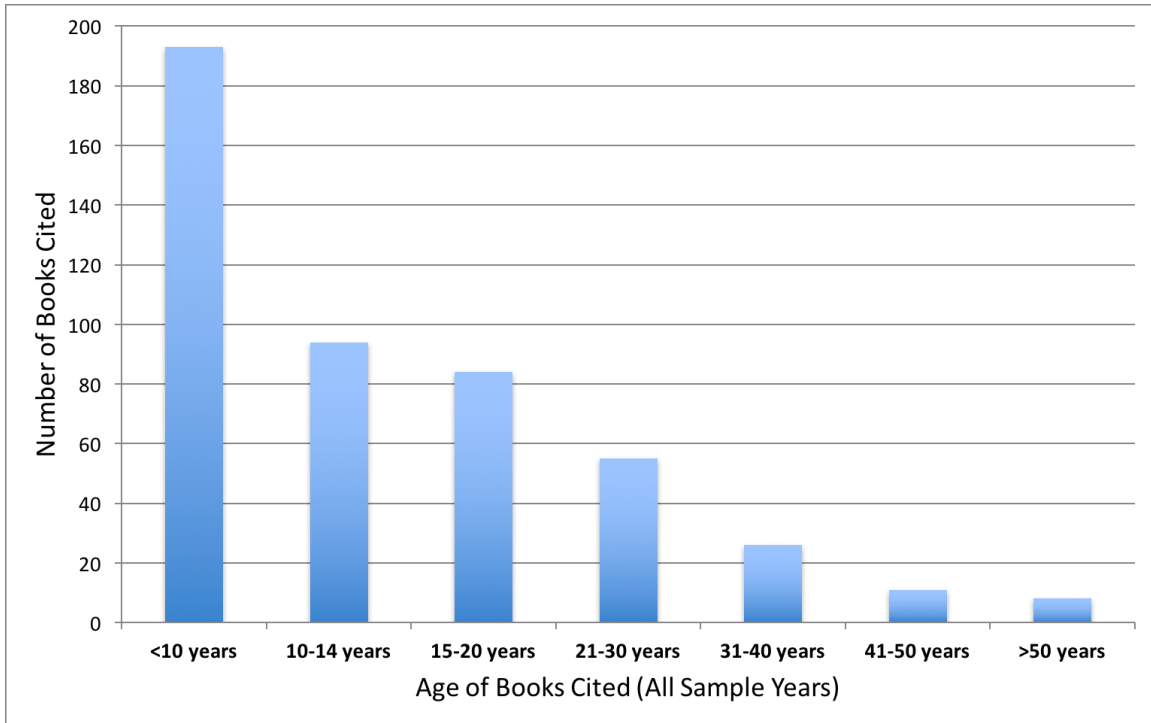
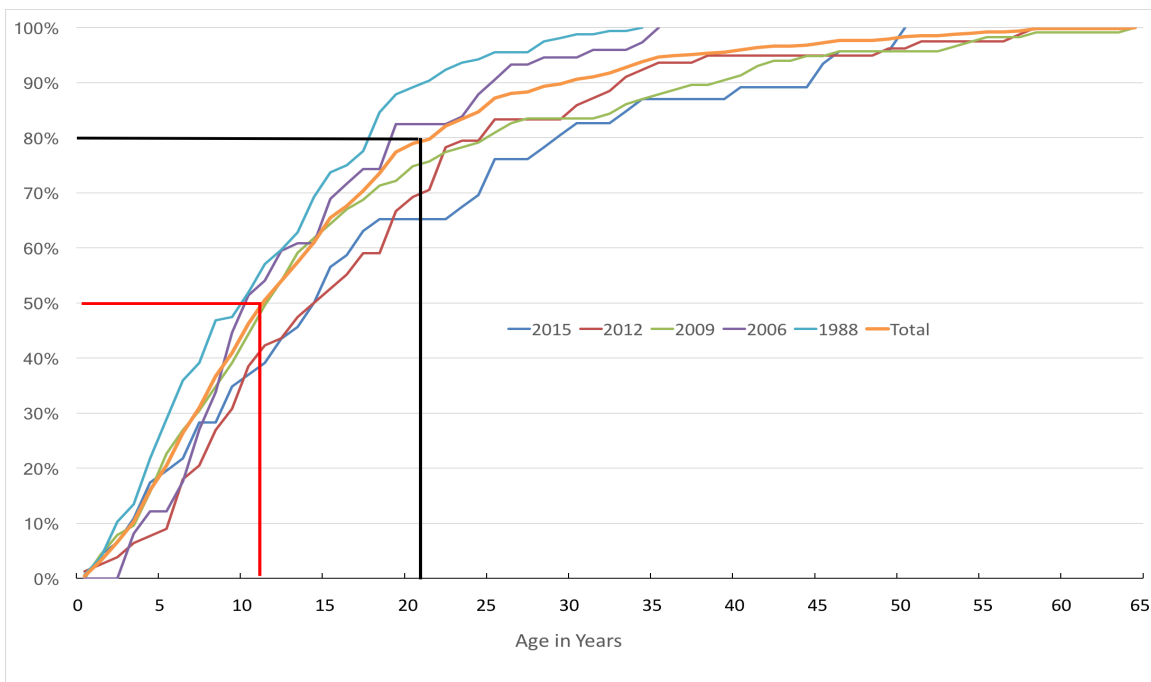


Figure 5. Cumulative Percentage of Cited Books by Age

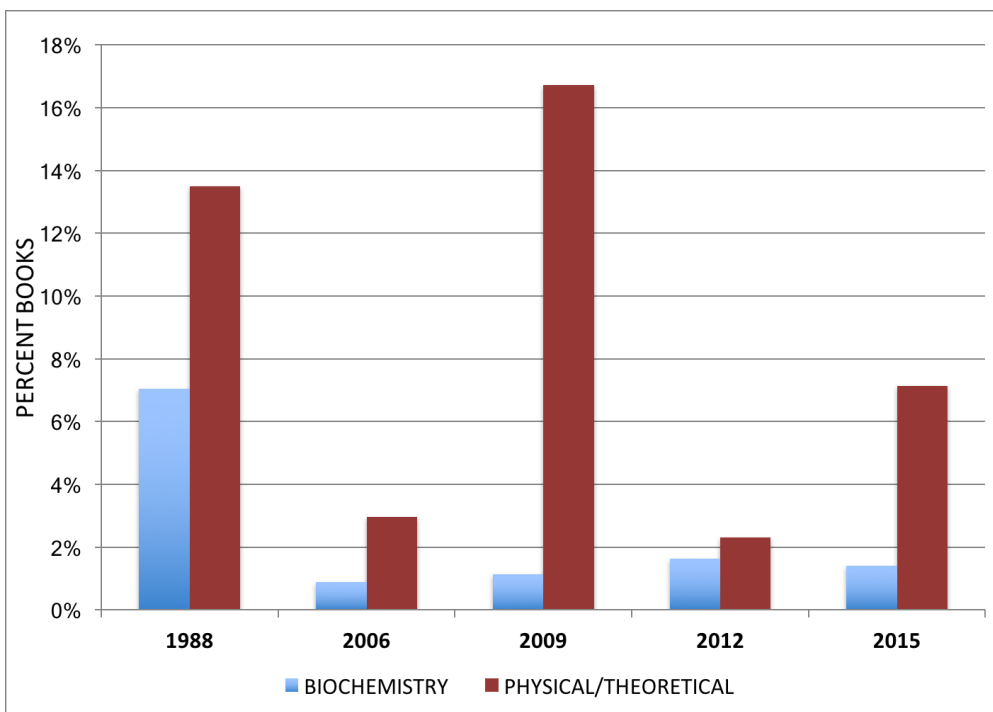


While each dissertation in the sample was assigned to one subject category, their distribution was not even: Biochemistry predominated with 15 in the sample, while Inorganic and Electrochemistry were the smallest categories (five and two respectively), as shown in Table 3. These proportions are roughly in line with the number of active research faculty in those specialties over time. The small sample size makes it difficult to reach general conclusions about book usage within these sub-disciplines, although it was apparent that dissertations in the Physical/Theoretical area tended to cite significantly more books than those in Biochemistry, the only two sub-groups represented in all five sample years (Figure 6).

Table 3. Number of Sampled Dissertations by Subject Category

<b>Subject Category</b>	<b>1988</b>	<b>2006</b>	<b>2009</b>	<b>2012</b>	<b>2015</b>	<b>Total</b>
Biochemistry	2	4	4	1	4	15
Organic	0	4	3	2	3	12
Physical/Theoretical	1	2	3	3	1	10
Analytical	3	0	0	3	0	6
Inorganic	3	0	0	1	1	5
Electrochemistry	1	0	0	0	1	2

Figure 6. Books as Percentage of Total Citations: Biochemistry and Physical/Theoretical Chemistry.



#### Comparison with Earlier Studies

This study found that books accounted for 5.4% of all citations in the 50 dissertations sampled (4.4% if the print 1988 baseline group is excluded). This figure fits in well with prior single-institution bibliometric studies of U.S. chemistry dissertations that had somewhat narrower time frames (Table 4). Interestingly, the non-U.S. analyses yielded significantly higher book percentages. None of these other studies, however, collected granular data on the characteristics of cited books, and none took a longitudinal approach to the question.

Table 4. Summary of Prior Citation Analyses of Chemistry Dissertations

<b>Paper</b>	<b>Years Studied</b>	<b>Pct. Books*</b>
Gohain, Saikia (Tezpur U., India, 2014)	2008-2012	15.6%
Vallmitjana, Sabaté (Institut Químic de Sarriá, Spain, 2008)	1995-2003	12%
Mubeen (Mangalore U., India, 1996)	1980-1993	11.48%
Gooden (Ohio State U., 2001)	1996-2000	8.4%
Zhang (Mississippi. State U., 2013)	2002-2011	7.1%
Present Study	1988-2015	5.4%
Kayongo, Helm (Notre Dame U., 2012)	2005-2007	3%

\* The number of significant digits reflects the data as reported in the original studies.

### The E-Book Question

The time span of ETDs selected for this study coincided with the advent of the electronic book format in academic libraries. In 2007 UT Austin embarked on an ambitious program of demand-driven acquisition (DDA) of vendor-supplied e-books, which ran alongside traditional print acquisition models until 2014, when it was ended due to budget constraints. Several non-DDA e-book plans operated simultaneously during this period. Discovery records for e-books regularly cycled in and out of the catalog during this time as library staff administered and adjusted the various profiles. No dissertations in the sample from any year explicitly cited an electronic book, which was expected since citation styles are intentionally format-agnostic. Therefore, it is not possible to speculate on how many, or if any, of the books cited were obtained or used by the authors as e-books. Nevertheless, one might reasonably expect to see a *rising* trend of book citation resulting from e-books' easier access (compared to print) as well as their increasing number over the period of study. In fact, the opposite was true, as book citations

declined during this time both in terms of absolute number and as a percentage of total citations (see Figures 1 and 2). Further research might shed more light on this paradox.

### The Ownership Dilemma

While the library owned the large majority (92%) of books cited in the sampled dissertations, the importance of local availability is a classic chicken-or-egg question: Do students use what the library owns, or does the library own what students need? Since UT Austin has large and deep science collections, it was expected that the library would own most of what they needed in the course of their research. Anecdotally, UT Austin chemistry graduate students are not heavy users of interlibrary loan for borrowing books not owned by the library, an indication that the library's collection is indeed meeting most of their book-type needs.

### Barriers to Citing Books

One obvious question emerges from these results: Why do chemistry graduate students cite so few books in their dissertations? We can propose some possible explanations, which are admittedly speculative.

- *Discovery challenges* – Books and book chapters tend not to be indexed in major databases (e.g. SciFinder, Web of Science) or cited in journal articles, and thus may go unnoticed in the course of a student's research.
- *Adviser/Peer bias* – Students may be discouraged, either actively or implicitly, by their mentors and peers from using and citing books, if the latter view books as a lesser form of scientific literature due to lack of peer review, lack of currency, or perceived lack of authority.
- *Disciplinary conventions* – Chemistry and Biochemistry, as mentioned earlier, are highly journal-centric disciplines where it is expected that journal articles will dominate any reference list. Dissertation bibliographies understandably reflect those of the literature they are based on – a self-perpetuating bias.
- *Physical access barriers* – In an era of digital instant gratification, obtaining a hardcopy book requires more effort and time even if one is aware of its existence and potential usefulness. The walk to the library, the hunt in the stacks, and wait

at the desk are non-trivial obstacles to busy students. The growing trend of storing books remotely and requiring patrons to submit a retrieval request and wait days for delivery prevents the rapid browsing of contents and further discourages their use.

- *Utility vs. citation* – Students may use books as helpful review sources of background knowledge and literature citations, but end up not citing them in the dissertation's literature review.

Are there ways that librarians can encourage graduate students to use (and cite) books more often? Libraries have made considerable financial and operational investments in their book collections, both in print and digital formats. It is reasonable to take actions that will increase the return on that investment. Some possible strategies might include:

- Improve discovery mechanisms. Library catalogs are rarely first-stop resources for STEM students, and links into local catalogs from other types of tools (such as Amazon, Google Books, index databases, etc.) are crude or nonexistent.
- Reduce barriers to access. Make it easier to locate, request, and check out physical books.
- Keep core print monographs close to users. Proximity matters, and remote location trends driven by the desire to conserve or repurpose library space certainly depress the subsequent use of book collections. (See Note 3) The age data shown in Figures 4 and 5 above suggest that voluntary book relocation strategies in chemistry should focus on titles more than 20 years old to conform to the 80/20 rule.
- Build awareness of the usefulness and purpose of books in information literacy curricula and outreach efforts aimed at graduate students.

## **Conclusion**

This study sought to assess the level of book usage among chemistry graduate students by analyzing the citation patterns to books in 50 randomly sampled PhD dissertations from



five completion years: ten from 1988 (pre-ETD baseline), and ten each from more recent three-year intervals: 2006, 2009, 2012, and 2015. Of a total of 8,723 references, 474 (5.4%) were coded as “books.” The median age of the books cited was 11 years, and 80% were 21 years old or less. 436 of the 474 books (92%) were judged to have been available in the university’s library collection at the time of completion. The level of e-book usage could not be assessed because the format of books cited is not typically indicated in a citation.

The trend of book citation in chemistry dissertations over the years studied was gradually but clearly downward. These data, coupled with longitudinal circulation data, indicate that graduate students’ use of the library’s print book collection is declining, but the precise reasons are unclear. While secondary literature in general, and books in particular, have long been of lesser importance to chemists relative to primary journals, further research could shed light on why usage of monographic materials, as reflected by citations, is falling. In order to maximize return on the considerable investment made in current and legacy book collections, libraries should consider strategies to increase use of books by working to reduce various barriers to discovery and access, making smarter, data-driven decisions about where book collections should be optimally stored, and increasing student awareness of books as a useful form of scientific literature.

## Notes

1. The Biochemistry Division was split off from the Department of Chemistry and Biochemistry in 2013 and merged into a newly formed Department of Molecular Biosciences encompassing biochemistry, genetics, and molecular biology (with over 60 faculty). The downsized Department of Chemistry was left with about 30 faculty.
2. The ostensible justification for the ETD embargo option, which is routinely recommended by chemistry faculty advisers and largely unique to that department, is to avoid potential interference with pending patent applications or journal article submissions. Ramirez et al. (2014) reported that, contrary to anecdotal perceptions, most STEM publishers and editors, including those in

chemistry, routinely accept submissions derived from ETDs. Journal editorial policies are often vague on this point, however -- a case in point being the highly variable [policies of journals published by the American Chemical Society](#).

3. This point has been convincingly demonstrated at UT Austin by two recent science branch closures. The Engineering Library closed and moved out of its space in 2013 due to a major building construction project, and the circulating books were moved several blocks away to the main library until a new engineering library opened in early 2018. Book circulation following the closure immediately dropped by half and stayed at that level thereafter. The Chemistry Library closed in May 2017 due to a multiyear building renovation project. Its circulating books were also relocated to the main library stacks, and circulation of that collection in the seven months that followed was 43% lower than the same period in 2016.

## References

Anonymous. 2016. The past, present and future of the PhD thesis. *Nature*. 365(7610): 7. DOI: 10.1038/535007a

Eckel, E. J. 2009. The emerging engineering scholar: A citation analysis of theses and dissertations at Western Michigan University. *Issues in Science and Technology Librarianship* 56. Available from: <http://www.istl.org/09-winter/refereed.html>

Edwards, S. 1999. Citation analysis as a collection development tool: a bibliometric study of polymer science theses and dissertations. *Serials Review*. 25 (1): 11–20. DOI: [https://doi.org/10.1016/S0098-7913\(99\)80133-6](https://doi.org/10.1016/S0098-7913(99)80133-6)

Gohain, A. & Saikia, M. 2014. Citation analysis of Ph.D theses submitted to the Department of Chemical Sciences, Tezpur University, Assam. *Library*

*Philosophy and Practice*. (Winter) 1066. Available from:  
<http://digitalcommons.unl.edu/libphilprac/1066>

Gooden, A. M. 2001. Citation analysis of chemistry doctoral dissertations: an Ohio State University case study. *Issues in Science and Technology Librarianship*. 32. Available from: <http://www.istl.org/01-fall/refereed.html>

Gould, Julie. 2016. What's the point of the PhD thesis? *Nature*. 535(7610): 26-28. DOI: 10.1038/535026a

Kayongo, J., & Helm, C. 2012. Relevance of library collections for graduate student research: a citation analysis study of doctoral dissertations at Notre Dame. *College & Research Libraries*. 73 (1): 47–67. DOI: <https://doi.org/10.5860/crl-211>

Kushkowski, J. D., Parsons, K. A. & Wiese, W.H. 2003. Master's and doctoral thesis citations: analysis and trends of a longitudinal study. *portal: Libraries and the Academy*. 3(3): 459–479. Available from: <https://muse.jhu.edu/article/45648>

Mubeen, M.A. 1996. Citation analysis of doctoral dissertations in chemistry. *Annals of Library Science and Documentation*. 42(3): 48-58. Available from: <http://nopr.niscair.res.in/handle/123456789/27538>

Ortega, L. 2008. Age of references in chemistry articles: A study of local authors' publications from selected years, 1975–2005. *Science & Technology Libraries*. 28 (3): 209–246. DOI: <https://doi.org/10.1300/01942620802098768>

Ramirez, M.L., McMillan, G., Dalton, J.T., Hanlon, A., Smith, H.S., & Kern, C. 2014. Do open access electronic theses and dissertations diminish publishing opportunities in the sciences? *College & Research Libraries*. 75(6): 808-821. DOI: <https://doi.org/10.5860/crl.75.6.808>

Subramanyam, K. 1980. Citation analysis in science and technology. In: R.D. Stueart and G. B. Miller, editors. *Collection development in libraries: a treatise*. Greenwich: JAI Press. p. 345–372.

Vallmitjana, N. & Sabaté, L.G. 2008. Citation analysis of Ph.D. dissertation references as a tool for collection management in an academic chemistry library. *College & Research Libraries*. 69(1): 72–82. DOI: <https://doi.org/10.5860/crl.69.1.72>

Zhang, Li. 2013. A Comparison of the citation patterns of doctoral students in chemistry versus chemical engineering at Mississippi State University, 2002–2011. *Science & Technology Libraries*. 32(3), 299-313. DOI: 10.1080/0194262X.2013.791169

[Disclaimer: The author, who was editor of the Refereed Articles section of this journal at the time of submission, was recused from the editorial process related to this article.]