

Comparing Characteristics of Highly Circulated Titles for Demand-Driven Collection Development

William A. Britten and Judith D. Webster

Collection practices in the current era of strict budgets are necessarily driven by the needs of the users. Use studies are therefore of increasing value to collection managers, but typically have provided only generalized statistical data. An alternative methodology is presented for analyzing the MARC records of highly circulating titles in order to document common characteristics that would be predictive of future use of additions to the collection. Items are evaluated for commonality of subject heading, author, language, and imprint date for selected Library of Congress classes.



Imagine interviewing library users at the circulation desk to discover their reasons for checking out each title. Was it the subject matter? The author? The currency of the information? If all of these responses were gathered and sorted, patterns of usage that would be of great value for collection management might emerge. For example, as we discovered at the University of Tennessee, Knoxville (UTK), books on the Vietnam War are among the most sought-after information in our library. However, interviewing enough patrons to generate substantial data is impractical, as is standardizing and entering their responses into a database for analysis. Why not let the books reveal the patterns of usage? The data stored in on-line systems can be used to *infer* the characteristics patrons are seeking.

The present article builds on the authors' previous research and discusses an analysis of a sampling of the most-circulated titles in the University of Tennessee, Knoxville collection.¹ The purpose

of the study is to analyze the machine-readable cataloging (MARC) records of "star performers" in each Library of Congress (LC) class to discover if patterns of commonality exist among high-use titles. Our hypothesis is that there will be common characteristics among high-use titles and that selectors can use these data as a component of a collection development plan to purchase titles that have a high likelihood of being used. Libraries with automated systems will find the methodology presented here particularly valuable.

COLLECTING FOR USE

Libraries have long built collections on the basis of potential use. Now, however, libraries must be more responsive to the immediate needs of their users. As the published output of each subject discipline increases, and library budgets remain stagnant or shrink, demand-based or use-based collection development becomes almost mandatory.² In this environment use studies are of increasing value.

William A. Britten is Automation Librarian and Judith D. Webster is Head of Acquisitions at the University of Tennessee, Knoxville Library, Knoxville, Tennessee 37996.

The literature on circulation studies is voluminous. The classic work of Herman Fussler and Julian Simon and the landmark Pittsburgh study suggested that a high percentage of library collections are unused.³ Other research has characterized use distribution over subject disciplines and compared in-house use to circulation.⁴ Articles that apply circulation statistics to collection management issues are uncommon.⁵ Collection-use studies are typically statistical summaries of circulation patterns for an entire collection or portions of a collection. These studies, however, usually do not examine the use of specific titles. The American Library Association's *Guide to the Evaluation of Library Collections* describes use studies only in terms of the ratio between the number of titles and the number of times those titles circulated.⁶ The present study, however, employs title-level analysis to characterize useful items, thus providing a more complete picture of user demands. This type of analysis is not concerned with absolute measures of use or nonuse, but seeks to determine which individual titles were useful.

ANALYZING ACTIVE TITLES

The University of Tennessee, Knoxville Libraries' collection includes more than 1.7 million items. The tracking of circulation activity for these items since 1982 has been accomplished by a Geac automated system. The portion of the collection included in this study is comprised of monographic items that have the highest cumulative circulation counts on the automated system (circulation data for periodicals is incomplete on the UTK automated system). After eliminating all but the monographic titles, there remained just less than 1 million items, which were first sorted by call number to group the LC classes. Sorting by class allowed data to be gathered that would be useful to subject bibliographers. Then the items were sorted in descending order by the total number of times circulated on the Geac system from 1982 through 1990. The result was a ranking of all titles in each LC class from most circulated to least circulated.

The next step was to select portions of the collection to examine. Since the plan was to inspect each title, it was decided that the study would be confined to the top 400 circulating titles of LC classes with at least 10,000 items. A group of 20 LC classes resulted, which represented large segments of the collection where significant amounts of money were spent and bibliographers had much selection work to do. Choosing to study the top 400 titles in each of these 20 groups was an arbitrary decision, but it allowed investigation of the cream of the crop—titles which ranged in number of circulations from a high of 161 for a title in class LB to a "low" of 8 circulations for a title in class D. These 8,000 titles (20 groups of 400) averaged over 26 circulations each (the average circulation rate for all 921,596 monographic items in the collection is 2.65 circulations per item).

Comparing the MARC records of thousands of titles for indications of commonality was potentially labor intensive and intimidating. The library's integrated system, however, allowed extracting specific MARC tags from the online catalog database for the records gathered in the circulation database. Utilizing this technique it was possible, for example, to capture all of the subject headings, all of the authors, or all of the dates of publication for any of the groups of 400 highly used titles. These groups of MARC tags were then sorted to allow easy visual inspection to spot the clusters of commonality that characterize popular titles.

THE COMMONALITY FACTORS

The 20 groups of 400 popular titles were first analyzed for frequency of subject heading occurrence. Table 1 shows the 10 most frequently occurring subject headings for 5 of the 20 LC classes in the study. Table 1 reveals that among the top 400 circulated books in class BF, 39 were assigned the subject heading Nonverbal Communication, while in class HV, 51 of the top items had the subject heading Child Abuse, and in class PT over 25% of the top 400 have Henrik Ibsen as the subject. We can also observe that circulation

TABLE 1
OCCURRENCES OF SUBJECT HEADINGS FROM THE
400 MOST-CIRCULATED BOOKS FOR SELECTED LC CLASSES

BF	Nonverbal communication	39
BF	Stress	37
BF	Cognition	21
BF	Dreams	20
BF	Psychoanalysis	16
BF	Attitude	14
BF	Witchcraft	14
BF	Anxiety	11
BF	Interpersonal relations	9
BF	Control	8
HV	Child abuse	51
HV	Capital punishment	50
HV	Suicide	36
HV	Alcoholics/alcoholism	27
HV	Rape	23
HV	Sign language/deaf, means of communication	19
HV	Animals, treatment of	18
HV	Wife abuse	17
HV	Children, deaf	15
HV	Deaf education	12
RC	Anorexia nervosa/bulimarexia	45
RC	Family psychotherapy	27
RC	AIDS	24
RC	Depression, mental	22
RC	Psychotherapy	21
RC	Schizophrenia	19
RC	Rational-emotive psychotherapy	17
RC	Psychoanalysis	14
RC	Obesity	12
RC	Alcoholism	11
PS	Frost, Robert	48
PS	Miller, Arthur	30
PS	O'Connor, Flannery	28
PS	Williams, Tennessee	27
PS	Hawthorne, Nathaniel	24
PS	Poe, Edgar Allen	23
PS	Plath, Sylvia	16
PS	Hemingway, Ernest	10
PS	Dickinson, Emily	8
PS	Hughs, Langston	8

TABLE 1 (continued)

PT	Ibsen, Henrik	108
PT	Kafka, Franz	46
PT	Goethe	15
PT	Strindberg, August	14
PT	Brecht, Bertold	12
PT	German literature—18th century	6
PT	Dramatists—Norwegian—biography	5
PT	Publishers and publishing—Germany	4
PT	Cotta, Johan	4
PT	Marat, Jean Paul	4

TABLE 2
DISTRIBUTION OF IMPRINT DATES FROM THE
400 MOST-CIRCULATED BOOKS FOR SELECTED LC CLASSES

	Pre-1960	1960s	1970s	1980s
B	78	116	127	78
BF	24	52	198	126
D	58	105	129	106
DA	86	80	143	89
DS	85	126	136	50
E	44	102	158	96
F	86	79	114	120
HC	19	58	141	182
HD	18	15	110	256
HF	18	44	173	165
HV	10	37	166	186
LB	10	47	181	162
PN	53	88	121	137
PQ	85	126	136	50
PR	65	154	116	63
PS	30	132	160	77
PT	86	132	119	56
QA	16	60	137	187
QC	44	101	136	119
RC	7	23	128	241

in class PS is dominated by criticism of poets and playwrights, while books related to deafness are very popular in class HV. Examination of a listing of the full records for the top items in each class further divulged that of the 25 most-

used titles in class HV, 18 had the subject heading Capital Punishment, while 13 of the top 25 titles in class RC were about anorexia nervosa.

The data set was next analyzed for imprint dates. Table 2 shows a breakdown

TABLE 3
 DISTRIBUTION OF BOOKS ABOUT AN AUTHOR VS. BOOKS BY AN AUTHOR
 FROM THE 400 MOST-CIRCULATED BOOKS FOR SELECTED LC CLASSES

LC Class	Author	Books about	Books by
PQ	Samuel Beckett	7	15
PQ	Albert Camus	24	11
PQ	Molière	18	12
PQ	Jean Paul Sartre	5	10
PR	William Shakespeare	46	15
PR	Geoffrey Chaucer	19	9
PS	Robert Frost	48	7
PS	Steven King	0	7
PS	Arthur Miller	30	2
PS	Sylvia Plath	16	8
PT	Goethe	15	22
PT	Henrik Ibsen	108	34
PT	Franz Kafka	46	8

of dates of publication for all 20 classes into the categories pre-1960, 1960s, 1970s, and 1980s.

Given that the data represents circulations recorded from 1982 to 1990, table 2 clearly indicates that titles remain well used for many years after their publication, even outside the areas where this might be expected, such as literature and history. In fact, it appears that for the UTK Libraries' monograph collection, the hard sciences (represented by physics—QC and math—QA) are less currency-oriented than many of the social sciences (economic history—HC, social welfare—HV). Class RC is the most contemporary-conscious, with only 30 books published prior to 1970 represented among the top 400 circulating items. The data presented in table 2 indicate that further study is needed before weeding older titles or instituting a collection policy that excludes the purchase of all but the most recently published titles.

Analyzing the occurrence of authors among the top items in each class proved to be useful only in the literature classes of the study—PQ, PR, PS, and PT. Among the nonfiction classes there was almost no commonality related to authors, and when an author did appear

several times in the top 400 list it was nearly always due to the library's owning multiple copies of one title. However, studying the occurrence of authors in literature allowed a comparison between the use of criticism versus the use of the original work. For example, table 3 shows a comparison of books about an author versus books by an author for books which appeared among the 400 most used titles of their respective classes. Curiously, while books of criticism are usually favored over the original works, the opposite is true for Beckett, Goethe, and Sartre.

Titles remain well-used for many years after their publication, even outside the areas where this might be expected.

Lastly, the language indicator in MARC field 008 was analyzed for LC classes PQ and PT. Table 4 confirms the expectation that foreign language titles do not circulate well. In fact, the most highly used titles in PT are in English. The overwhelming lack of highly circulating titles in the German language in

TABLE 4
LANGUAGE COMPOSITION OF SELECTED LC CLASSES VS.
LANGUAGE COMPOSITION OF TOP 400 CIRCULATING BOOKS

Language	% of Language in Class	% of Language in Top 400
PQ		
English	27	79
French	32	6
Italian	6	0
Portuguese	2	1
Spanish	31	14
PT		
English	17	83
German	80	16

TABLE 5
CIRCULATION RATES FOR SELECTED FOREIGN LANGUAGES
(MONOGRAPHS ONLY—FOR SELECTED LC CLASSES)

Language		No. of Books	Circs. per Book
Chinese	(PL)	204	2.51
Japanese	(PL)	598	.82
Spanish	(PQ)	10,226	.78
French	(PQ)	10,392	.74
Portuguese	(PQ)	789	.67
German	(PT)	17,286	.33
Italian	(PQ)	1,843	.22
Russian	(PG)	5,807	.14
All monographs		921,596	2.65

class PT is surprising, however, considering that 80% of PT purchases are in German. Conversely, the relatively small percentage of Spanish titles in PQ accounts for a fairly high number of the high-use items.

Given this indication of foreign title use, it seemed appropriate to evaluate the circulation rates for all titles in these two classes. Table 5 shows the average number of circulations per book for the major languages in classes PQ and PT, as well as for classes PL and PG, two classes that were not in the overall study.

Even though circulation rates are often dismal, research libraries cannot cease buying foreign language titles. To ensure a higher probability of buying

useful items, however, the methodology presented in this study could be applied separately to the top-circulating foreign titles to determine their common characteristics.

PEER COMPARISONS

Table 1 includes some of the most popular subject headings in the UTK collection. The holdings of other research libraries were examined for several of these subjects to assess the comparative strength of the UTK collection. Three peer libraries were chosen using these criteria: similarity of collection size, ability to perform keyword searches in the online catalog, and remote Internet accessibility.⁷ Since the UTK collection is fully

converted to machine-readable records, items located in peer online catalogs but not found in the UTK online catalog would be potential candidates for collecting at UTK. Keyword searches were performed using the online catalogs of the University of Iowa, the University of Minnesota, and Michigan State University for several terms, including *nonverbal communication* and *anorexia*.

A keyword search of the three peer libraries' online catalogs for the term *nonverbal communication* yielded similar numbers from citations for the 4 libraries (UTK 201, Iowa 207, Michigan State 226, and Minnesota 247). However, the evidence suggests that the collection at UTK related to this topic needs additional titles or duplicates: 99% of all items with the subject heading *Nonverbal Communication* have circulated, with an average of 24 circulations per book.

Searching the 3 remote catalogs for books related to anorexia produced striking results: UTK 45 citations, Michigan State 85, Iowa 90, and Minnesota 124. Interestingly, all 45 of UTK's items on anorexia appear in the list of the top 400 circulating books for class RC. Since it appears that UTK has undercollected for this subject, the next step was to download the citations from the three peer catalogs and compare them with those held by UTK. The screen-capturing utility SCAP was used to capture brief title/date citations from the four catalogs. The file was then imported into a spreadsheet and sorted by title so that it would be easy to note which titles appeared in more than one catalog. The resulting "union list" spawned a list of over 50 potential additions to the UTK collection, including 8 titles published since 1985 and held by all 3 peer libraries, and 15 additional titles published since 1985 and held by 2 out of 3 peers.

Remote OPAC comparisons produced similar results for the subjects *capital punishment* and *Stonehenge*. Titles relating to capital punishment accounted for 18 of the top 25 circulating books in class HV, and the analysis of peer library holdings revealed that UTK's collection had fewer titles in this area.

The subject of *Stonehenge* was something of an oddity: there are only 10 items in the UTK collection in class DA (British History) with the subject heading *Stonehenge*. These 10 books are extremely well used—circulating an average of 35 times each, with 7 of them among the top 10 circulating items for DA. A check of the peer libraries located 14 additional titles, with only 4 held by all 3 peers. Searches of the online catalogs of several additional large research libraries established that there simply is not much to be collected on *Stonehenge*.

A tool such as the OCLC/Amigos Collection Analysis System on CD-ROM could be used as an alternative to comparisons made from searching remote online catalogs. The OCLC/Amigos system was not available for this study, but it appears to be an ideal complement to this type of collection use analysis.⁸

THE UNUSUAL CASE OF PT

Class PT emerged during the study as a highly unusual collection. The popular subject headings listed in table 1 for PT indicate that criticism about a few authors is highly used, while table 4 shows that the collection is primarily comprised of titles in German that do not circulate well. Further investigation was clearly needed.

The Germanic literatures collection (PT) includes over 24,000 monographic items. The collection circulates at a rate of .69 per item, compared to a rate of 2.65 circulations per item for the entire UTK monographic collection. PT ranks at the low end of the relative circulation rate scale, with classes BF (6.35 circulations per item), HV (5.25), and RC (5.13) near the top. Furthermore, only 22% of PT monographs have ever circulated on the automated system during the 1982-90 period, compared with 77% for BF, 75% for HV, and 76% for RC. Finally, only 17% of the books in PT are in the English language, yet they account for over 60% of the class's circulations. The statistics indicate that PT has problems, but they do not provide title-level details. The methodology presented in this study reveals characteristics of the PT titles that *have* circulated.

Table 1 indicates that books about Henrik Ibsen are very much sought after in class PT. In fact, of the top 25 circulating PT items, 22 are either about Ibsen or were written by Ibsen. Investigating further, we find that there are 173 books about Ibsen or written by Ibsen in class PT, with more than 89% of them circulating during the 1982-90 period, and averaging over 11 circulations per item. The 173 Ibsen-related items represent less than 1% of the PT collection, yet they account for over 18% of all PT circulations. Moreover, only two of the Ibsen-related books are in the German language and neither has circulated. Similar statistics exist to a lesser degree for the other top authors of class PT.

A keyword search of Ibsen in the online catalogs of the 3 peer libraries reveals that while the UTK collection has 215 Ibsen-related items (some are not in PT), Michigan State has 319, Iowa 402, and Minnesota 723. Clearly the Ibsen collection at UTK needs to be expanded.

Overall, the PT collection is comprised of a preponderance of German-language titles that are little used, while what does circulate is overwhelmingly skewed towards a select few literary names and books published in the English language.

FOLLOWING UP

The information generated by the methodology presented here often suggests areas for further investigation. Several examples of such follow-up activities are presented here.

The data uncovered many subject headings that were extremely popular, as indicated by high circulation levels. But were these topics consistently popular over time or had they been hot topics for a year or two? The answer to this question would be important when deciding to add titles to the collection. Since the UTK online system does not retain the details of circulation transactions, date due slips of books were examined for 3 subjects discussed previously: anorexia, Stonehenge, and nonverbal communication. Checking date due slips for these 3 topics showed evidence of consistent demand.

Another follow-up activity for subjects that appear to be overburdened is an analysis of titles collected by peer libraries. Aside from simply adding these titles to our collection, we attempted to determine what selection strategies would be likely to locate such titles in the future. For example, among the popular titles in BF held by the 3 peer libraries, it was noted that several were from one publisher (this same publisher showed up in interlibrary loan borrowing reports), several were published in Britain, and several were numbered series. Obtaining the publisher's catalog, locating more British sources, and subscribing to 1 or several of the series would be ways to ensure higher levels of future collection in these subject areas.

LC class QC (Physics) was one of the 20 classes included in the study. The distribution of imprint dates (table 2) showed that a relatively high number of the top circulating books were published prior to 1970. Because currency of information is assumed to be of importance for science titles, further investigation was done. Of the top 400 circulating books in QC, 51 of them have the subject heading *quantum theory*. Looking at the full records of these 51 books, it was noted that the majority of them were basic texts, as evidenced by titles beginning *Introduction to . . .*, *Elementary Quantum . . .*, and *Principles of . . .*. Also, the majority were older texts. These facts prompted a follow-up extract of all books in QC with the subject heading *quantum theory*, which indicated that all basic titles on quantum theory are very much in demand, but the collection does not contain many recent publications. Again, Internet searching of peer library catalogs was used to locate candidates to fill this gap.

CONCLUSION

The methodology offered here is a valid means of assessing trends of demand for specific types of items in a library's collection, for uncovering areas that have been undercollected and are overburdened with use, and for exposing areas that have been well collected but rarely circulate. As part of an overall

collection management program, the data should be interpreted within the context of the user environment—the curriculum, faculty research interests, etc. Although the study was conducted with data collected over an 8-year period, analyzing consecutive periods of shorter duration would establish the staying power of popular subjects, resolving the problem of checking date-due slips.

The method presented here provides practical techniques that can be replicated in libraries with automated systems.

Use and user studies have a long history and remain important means of evaluating library collections and determining future directions for collection development. Traditional methods of studying use, however, often involve an unreasonable employment of librarians' time.⁹ The method presented here provides practical techniques that can be replicated in libraries with automated systems. The methodology rests on the ability to create a database of MARC records sorted by call number and including circulation counts. Once this is done, subsequent extracts of subject headings, authors, imprint dates, etc. for a portion of the collection are easily obtained. An LC class can be quickly analyzed at the request of a selector.

Most use studies verify what many librarians suspect: that a small percentage of a collection accounts for a large percentage of the circulations.¹⁰ The methodology presented here will assist librarians in selecting titles that will be

used. These strategies illustrate a beneficial partnership between collection development librarians and automation librarians. The automation librarian extracts and manipulates use statistics in a variety of ways, while collection development librarians interpret the data and incorporate them as selection tools. The partnership will result in a collection that reflects the needs of the local user population.

Would libraries following this methodology be sacrificing a collection broadly representative of all publications in all subjects? Yes, this method advocates giving precedence to items in categories of known popularity over those which have attracted little use over time. Would the collection eventually stagnate as users are offered only items which had been used before? No, our method is meant to be a component of an overall collection-development plan that would also include traditional methods of selection—it is the *emphasis* that is shifting. If all libraries adopted the practice of comparing holdings, would that foster homogenization, with all collections tending toward a similar core? Not at all! Following this methodology would lead to a collection that reflects user demand, which in turn reflects the unique characteristics of each library's constituency.

By collecting more on the basis of anticipated demand and minimizing the purchase of items known to circulate infrequently, libraries may find that they can satisfy user demands in spite of shrinking budgets. Along with other collection-use statistics, the reports presented in this study represent the tools of the trade for demand-driven collection development in an automated environment.

REFERENCES

1. William A. Britten and Judith D. Webster, "Class Relationships: Circulation Data, Collection Development Priorities, and Funding for the Future," *The Bottom Line* 4:8-11 (Spring 1990).
2. Ross Atkinson, "Old Forms, New Forms: The Challenge of Collection Development," *College & Research Libraries* 50:507-20 (Sept. 1989).
3. Herman Fussler and Julian Simon, *Patterns in the Use of Books in Large Research Libraries* (Chicago: Univ. of Chicago Pr., 1969); Allen Kent and others, *Use of Library Materials: The University of Pittsburgh Study*, (New York: Dekker, 1979).

4. Paul Metz and Charles Litchfield, "Measuring Collections Use at Virginia Tech," *College & Research Libraries* 49:501-13 (Nov. 1988); George M. Jenks, "Circulation and Its Relationship to the Book Collection and Academic Departments," *College & Research Libraries* 37:145-53 (Mar. 1976); Anthony Hindle and Michael K. Buckland, "In-Library Book Usage in Relation to Circulation," *Collection Management* 2:265-77 (Winter 1978).
5. Britten and Webster, "Class Relationships"; Adrian N. Peasegood, "Towards Demanded Book Acquisitions?: Experiences in the University of Sussex Library," *Journal of Librarianship* 18:242-56 (Oct. 1986).
6. Barbara Lockett, ed., *Guide to the Evaluation of Library Collections* (Chicago: American Library Assn., 1989), p.9.
7. William A. Britten, "BITNET and the Internet: Scholarly Networks for Librarians," *College & Research Libraries News* 51:103-7 (Feb. 1990).
8. "Collection Analysis Compact Disc System," *Technical Services Quarterly* 7:67-68 (1990).
9. Robert W. Burns, "Library Use as a Performance Measure: Its Background and Rationale," *Journal of Academic Librarianship* 4:4-11 (Mar. 1978).
10. R. W. Trueswell, "Some Behavioral Patterns of Library Users: The 80/20 Rule," *Wilson Library Bulletin* 43:458-61 (Jan. 1969); William A. Britten, "A Use Statistic for Collection Management: The 80/20 Rule Revisited," *Library Acquisitions: Practice and Theory* 14:183-89 (1990).

NEW PUBLICATIONS FROM ACRL

ALA Order Dept. 1-800-545-2433, 50 E. Huron St, Chicago, IL 60611-2795

▼ ACRL/Historically Black

**Colleges & Universities
Library Statistics, 1988-89**
\$35.95/ACRL member \$25.95
ISBN 0-8389-7547-X

▼ Genre Terms: A Thesaurus for Use in Rare Book and Special Collections

Cataloguing (2nd ed.)
\$19.95/ACRL member \$16.95
ISBN 0-8389-7516-X

▼ *College & Research Libraries* and *College & Research Libraries News*, Index for Volumes 41-50 (1980-1989)

\$29.95/ACRL member \$25.95
ISBN 0-8389-7487-2

▼ Read This First: An Owner's Guide to the New Model Statement of Objectives for Academic Bibliographic Instruction

\$17.95/ACRL member \$14.95
ISBN 0-8389-7548-8

▼ Recruiting the Academic Library Director (ACRL) and The Search Committee Handbook (AAHE)

\$16.95/ACRL member \$13.95
ISBN (set) 0-8389-7484-8