Does the Catalog Record Make a Difference? Access Points and Book Use

Gunnar Knutson

This study builds on limited past research on correlations between catalog access points and circulation. The study examines 1,105 catalog records in three LC classifications and one subclass, comparing number and type of access points with number of circulations by type of user. Results show little significant positive correlation between total numbers of subjects or other access points and total circulation, but some interesting trends occur in correlations by type of subject and user. As we consider modifying or enhancing the catalog record, we must take into account such differences in users and subject areas.

s libraries make the transition from manual to automated systems, more information about the effects of such changes is needed in order to make best use of the new technologies. One area where more knowledge is required is in the relationship between cataloging and user behavior. We need to understand better the results of our past efforts to provide bibliographic access to academic library users. The current study provides further information on the question of whether circulation of library materials can be associated with elements of catalog records.

The traditional cataloging format, as exemplified by the Library of Congress–style catalog record, has served the library community well for many years and continues to answer different demands with its variety of access points. However, we should not consider ourselves bound by the limitations of the card catalog as we design new online systems. We need to test past assumptions in many small ways in order to develop a better composite picture of how users approach the catalog and how the catalog record succeeds or fails in meeting their needs.

The central concern of this study is the relationship between access points on catalog records and circulation: specifically, is there a correlation between an increased number of subject headings, and/or other access points, and the number of recorded uses of the book described? This is an attempt to discover whether there is a simple, quantitative approach to the problem of improving bibliographic access. If such a key exists, catalogers would be advised to increase numbers of access points, or at least certain types of access points such as subjects. In the course of trying to answer this question some other aspects of the bibliographic record that might affect circulation are also considered.

PAST RESEARCH

There is surprisingly little information linking catalog records and circulation.

Margaret Taylor concluded after a literature search for her 1982 doctoral dissertation that "there appears to be very little research which truly relates the idea of bibliographic accessibility, or the provision of bibliographic information, to the *use* of materials."[1] Studies that have addressed the question directly have all concluded with calls for more research, as there are many ways of approaching the problem. Taylor studied circulation patterns in a public library using the Dewey Decimal system. Pal V. Rao confined his more limited study (100 records) to the Library of Congress education (*L*) classification section of an academic library.[2] An earlier study by William Carl Highfill was conducted in a college library using the Dewey system.[3] It examined subject access points only, and was done during the summer session when use patterns may have been atypical.

While Rao and Taylor found essentially no significant correlation between number of access points and circulation, Highfill did conclude that "those books which have been assigned a greater number of subject headings have a greater chance of being selected by catalog users during subject searches."[4] Since type of library and type of collection, as well as the subject areas chosen for study, might have influenced these different findings, all of the authors called for further inquiry.

METHODOLOGY

The present study was conducted at a medium-sized research library that uses the Library of Congress classification system. The library's cataloged collection of about 800,000 volumes is housed on open shelves. A divided author/title and subject card catalog provides primary access to the collection; this is augmented by an automated circulation system with primary author and title search keys, but no added entries or subjects. There are no figures available on comparative use, but observation indicates that the card catalog is the primary source of bibliographic information.

A sample of one year's acquisitions in three different subject areas was studied and compared to circulation statistics for one fiscal year. The six months that elapsed between the date of the last acquisition and the beginning of circulation statistics was chosen to make certain that all books had been returned from the bindery and that all catalog cards had been filed.

The samples were taken from new book lists for the year 1981. Using a standard sample size formula provided by Taro Yamane,[5] systematic samples were drawn in three subject areas: law (class *K*, 232 records), fine arts (class *N*, 301 records), and social science (class *H*, 372 records). In addition, 200 records in the criminology section of social science (*HV*6001-9920, hereafter referred to simply as *HV*) were examined. (There was an overlap of 46 records that appeared in both the *H* and the *HV* samples.)

Three broad classifications were chosen to check if separate patterns in correlations could be observed. By choosing these distinct classifications, all of which exhibited a fair degree of circulation activity, the study was able to look for both broad and unique characteristics. The criminology group (*HV*) was added to provide an in-depth analysis of all 1981 acquisitions in an area with exceptionally high circulation—about twice that of the other classes in the sample.

For each record data were collected on the number and type of access points, plus date of publication, length, and language. To these were added circulation data derived from the library's automated circulation system for the fiscal year 1982–83. These data were entered by type of borrower: academic, graduate, undergraduate, and total. (The "academic" category includes all faculty, plus a small number of employees with faculty equivalent rank; the "total" category includes use by a small number of nonacademic, nonstudent employees.) Books that had circulated to nonuniversity users through interlibrary loan, etc., or had been placed on reserve during the year were removed from the study, because these circulations would probably not have occurred through use of the card catalog. Multicopy and multivolume sets were also removed because they had a greater potential for circulation. Only circulating items kept on

open shelves in the main library were considered for the sample.

All data were entered into a Time Zero Software statistical program designed for the Macintosh computer.[6] Each research hypothesis was that a significant positive correlation existed between the number of subjects or other access points and the total number of circulations during the year. Each null hypothesis being tested was that there was no significant correlation between the number of subjects or other access points and the number of times the item had circulated during the year. A 95 percent confidence interval was set for the tests. No cause-effect relationship was posited; what was sought was a level of correlation between the two factors that could not be explained by chance.

It is important to stress that what was being measured was not the relative circulation of one classification versus another, which would have been influenced by such factors as total size of the subject areas and university enrollment. The focus of the study was to examine the individual samples to see if a simple quantitative key could be found indicating that more of any of the various types of access points could be associated with significantly higher circulation.

Some limitations affecting the study must be acknowledged. Open shelving means that browsing will be a factor of unknown dimensions in circulation figures. Recent studies indicate that catalog users often approach the subject catalog as a starting point to guide them to a shelf area where they can browse for books they need,[7] and one portion of this study—class HV—seems to support this belief. The major reason for choosing the K classification as an area of study was to try to offset the browsing factor, as this classification is particularly difficult to browse. The size of the classified collection in the study library (around 800,000 volumes) would also tend to make browsing more difficult. Online author/title access to a majority of the collection through the automated circulation system surely had some effect on the non-subject statistics gathered, but would be of use only in known-item searching. This is discussed further in that section of the analysis.

FINDINGS

What follows is a summary of major findings and an attempt to apply such information to cataloging theory. Full statistical data are given in accompanying tables 1 and 2.

A central point of the study was to learn if books with more LC subject headings circulated more than similar books with fewer such headings. Subjects are the one area of the standard cataloging record that seem to offer the possibility of easy expansion. Most of the records in this study were unaltered MARC cataloging (more than 90 percent in each classification, and 95 percent total). The range of subject entries observed was between one and eight, with an overall average of 2.2 subjects per record. This is considerably higher than the average figures of 1.3 reported by McClure in 1976[8] and 1.4 reported by O'Neill and Aluri in 1981,[9] and can be attributed to the choice of nonliterary classes with recent imprint dates.

Classes H, K, and N have low positive total subject correlations with total circulation ranging from .046 to .088. HV has a negative correlation of − .072. In all cases the values are such that the negative hypothesis of no association cannot be rejected. That is, the figures are low enough that it cannot be stated that they could not have occurred by chance. The average number of LC subject headings per record, by itself, is not significantly associated with higher circulation in the four areas studied. An interesting finding is that undergraduate circulation figures are exceptionally high in the criminology section, HV. This, coupled with an accompanying negative total subject-circulation correlation, makes it seem possible that browsing plays a larger role in high-circulation areas, while the subject catalog is more important in less-used areas of the collection. For high-circulation areas the catalog may serve more as a guide to the general classification number than to a specific work.

Subject-heading data were recorded in two subcategories as well as a combined figure. The subcategories were labeled topical and specific.[10] "Specific" headings were any subject headings beginning with

TABLE 1

DESCRIPTIVE STATISTICS

Variable (average)	K (Law) 232 records	N (Fine Arts) 301 records	H (Social Sci.) 372 records	HV6001-9920 (Criminology) 200 records
Year of publication	1978	1977	1979	1979
Number of pages	312	220	264	251
English	96%	92%	99%	96%
Foreign	4%	8%	1%	4%
Topical subject	1.80	1.61	1.88	1.97
Specific subject	.25	.56	.38	.17
Total subject	2.05	2.17	2.25	2.14
Personal author	1.32	1.41	1.30	1.27
Corporate entries	.18	.23	.23	.17
Series entries	.47	.22	.37	.29
Title entries	1.02	.98	1.01	1.01
Total nonsubject	2.98	2.83	2.91	2.73
Total access points	5.03	5.00	5.16	4.87
Average number circulations				
Academic	.04	.07	.10	.12
Graduate	.15	.14	.24	.23
Undergraduate	.50	.61	.47	1.53
Total*	.70	.89	.85	1.92

*Includes local nonacademic circulation

TABLE 2

CORRELATION COEFFICIENTS

Correlation coefficients of *total* circulation and variables.
(Underlined values have significant correlation at .05 level)

Variable:	K (Law) 232 records	N (Fine Arts) 301 records	H (Social Sci.) 372 records	HV6001-9920 (Criminology) 200 records
Date	.068	.077	.041	.153
Length	−.015	.024	.137	.095
Language	.104	.093	.055	.158
Topical subject	.050	.042	.158	.020
Specific subject	.067	.006	−.116	−.198
Total subject	.088	.046	.071	−.072
Personal author	.016	−.007	.084	.050
Corporate entries	−.097	.078	−.094	−.107
Series	−.026	−.026	−.155	−.061
Titles	.066	.013	−.021	.077
Total nonsubjects	−.042	.031	−.076	−.051
Total access points	.028	.054	−.002	−.093

a personal or corporate name, or with a geographic name. One might expect that such headings would be easier for patrons to locate in the subject catalog and therefore might be correlated with more circulations. Other subjects—the "topical," or general, headings—are not always represented in the LC subject vocabulary in the same terms that may occur to a patron, and so conceivably might be associated with a lower correlation. Research shows that patrons rarely consult *Library of Congress Subject Headings* volumes for assis-tance in defining a subject.[11] As an example, a patron seeking a book on the art of Marcel Duchamp could locate it directly by a subject search under *Duchamp*, but might have trouble seeking a work on modern French art, which could be sought by such terms as *French art; modern art, French modern art*, etc., rather than the LC form "Art, Modern—20th century—France."

Just the opposite interpretation is supported by the total circulation figures. In three of the four total circulation catego-

ries, topical subject headings were more strongly correlated with circulation than were specific subject headings. There may be other factors that account for the differences—for instance, the type of books represented by these headings might be less in demand for other reasons.

Nevertheless, the aggregate figures do not support the idea that users are more likely to select a subject heading beginning with a proper noun.

When these statistics are broken down between types of users (see figures 1 and 2 and table 3) we see that there is a notice-
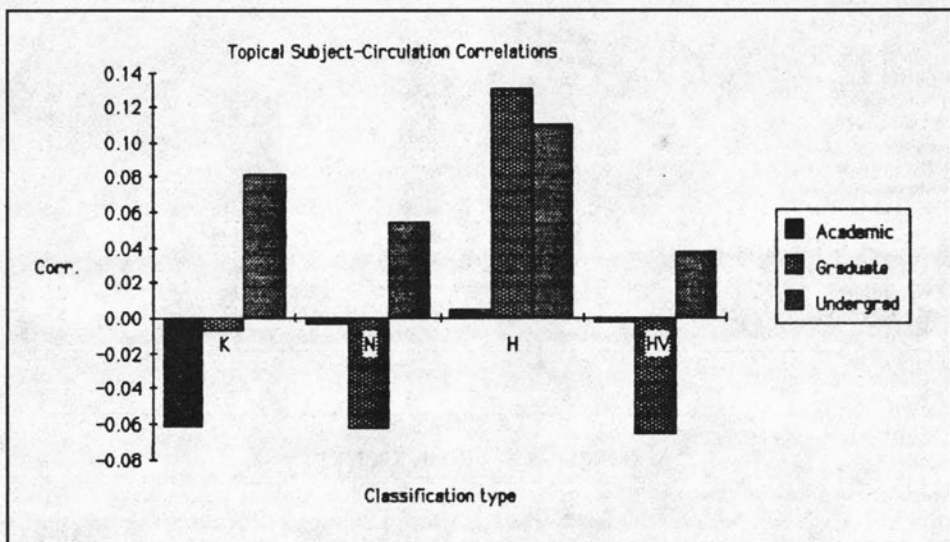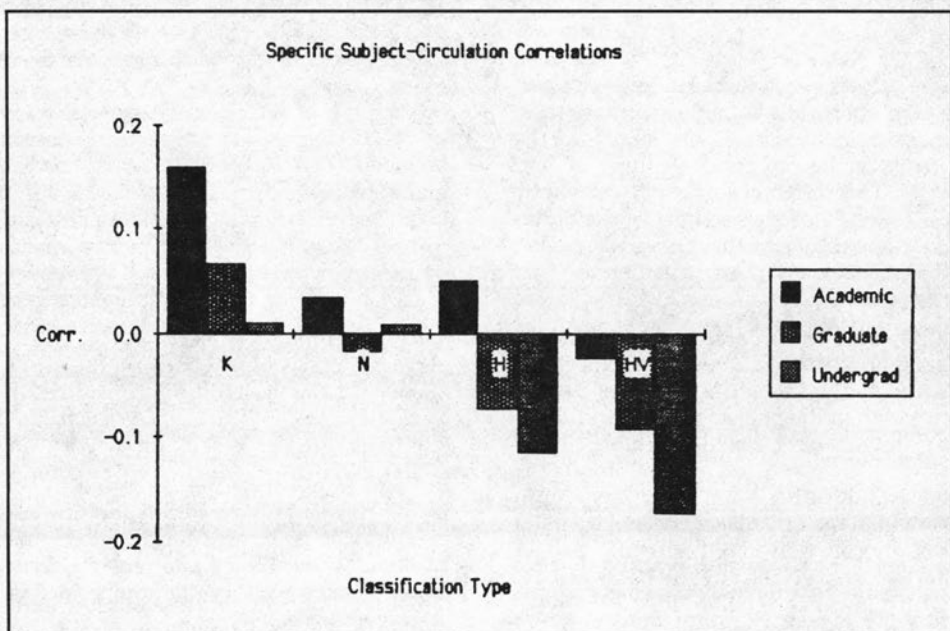

FIGURE 1

FIGURE 2

TABLE 3
CORRELATION COEFFICIENTS BY TYPE OF SUBJECT AND USER

(Underlined values have significant correlation at .05 level)

	K (Law) 232 records	N (Fine Arts) 301 records	H (Social Sci.) 372 records	HV6001-9920 (Criminology) 200 records
Topial Subjects:				
User group				
Academic	− .062	− .004	.004	− .003
Graduate	− .008	− .063	.130	− .067
Undergraduate	.082	.054	.110	.037
Total*	.050	.042	.158	.020
Specific Subjects:				
User group				
Academic	.159	.035	.050	− .025
Graduate	.067	− .018	− .072	− .094
Undergraduate	.010	.007	− .115	− .175
Total*	.067	.006	− .116	− .198
Total Subjects:				
User group				
Academic	.045	.018	.039	− .015
Graduate	.036	− .075	.075	− .110
Undergraduate	.081	.059	.024	− .043
Total*	.088	.046	.071	− .072

*Total figures include local nonacademic circulation

able difference in the topical and specific subject correlations between undergraduates and other users. Undergraduate use correlates positively to topical, or general, subject headings, while the academic and graduate use correlation figures are mostly negative. With specific subjects, i.e., subject headings beginning with personal, corporate, or geographic names, the undergraduate use correlations are the lowest of the categories in three of four cases. This difference might be caused by the type of materials sought by the different user groups, but that possibility is beyond the scope of the data gathered. What can be stated is that subject use to retrieve library materials is even more complex than indicated by previous studies.

Other access points are derived chiefly from the AACR2 rules by which a book is being cataloged. Although AACR2 gives some leeway in determining which access points to give a particular book, generally this is a standardized process. Depending on its descriptive cataloging, one work will receive only an author and a title entry, while another will also have access points for a series, joint author, and a prominent subtitle. It seems plausible that a patron approaching the catalog with an incomplete citation might have a slightly better chance of locating the book with more access points. Past research has shown that incomplete or garbled citations are common with library users,[12] and so it is possible that books with these extra access points might be charged out more often.

For three of the four classes examined (K, H, HV) there was a negative correlation between total nonsubject access points and total circulation. Class N had a slight positive correlation of .031. None of these figures disproved the negative hypothesis of no significant correlation. It is interesting to note that for nonsubject access points the most negative statistics came from series and corporate author correlations, while three-fourths of both the personal author and title correlations were positive. This suggests the possibility that patrons access materials infrequently through series or corporate authors. The higher correlations for author and title entries may be accounted for in part by the existence of the library's automated circulation system with its title and principal author search keys.

Access points are only part of the reason a patron might locate and choose a partic-

ular book. Some other factors were noted and analyzed. Date of publication, number of pages, and language were recorded as possible contributing factors. Correlation results were mixed but three noteworthy figures were seen. Date of publication did not vary widely because the sample was composed of all new additions from the same calendar year, with average dates of 1977 to 1979. Circulation correlations were positive in all cases (i.e., newer books circulated more often). With class *HV* the figure of .153 was high enough so that the null hypothesis of no significant association could be rejected. As one might expect, there was also a positive correlation between English language and circulation, and again the *HV* class figure of .158 was beyond the range of the null hypothesis. These figures would probably have been higher if there had been more foreign language books in the sample, but they ranged from only 1 to 8 percent of the totals. With such a preponderance of English works there was not much opportunity to observe the language factor.

Length is a more problematical factor as it is not easily associated with positive or negative characteristics. The correlations between greater length and higher use ranged from negative to positive, with the *H* class figure of .137 being high enough to indicate a significant positive correlation.

CIRCULATING BOOKS AS A SUBCLASS

A further set of tests was run following the conclusion of the initial phase of this study. As mentioned above, class *K* seemed the area most difficult to browse because of a mixture of local and LC classification. If any area of the study was most immune from the influence of browsing it would be the law classification. Approximately 33 percent of the 1981 acquisitions in this classification circulated during the 1982–83 fiscal year. (This compares to 39 percent in class *N*, 37 percent in class *H*, and 65 percent in class *HV*.)

There is some question of whether one can generalize about book use based on statistics that combine circulating and noncirculating books. It is possible that

items accounting for more circulation represent a different type of material than books that are seldom or never checked out. If the average noncirculating book has a distinctly different profile than the average circulating book it may be that factors such as publication date, length, and language, not numbers of access points, determine relative use. For such reasons it seemed desirable to break one class down into these component parts for further comparison.

The comparative figures are summarized in table 4. Circulating books as a group are somewhat newer (1979 as opposed to 1977 for noncirculating books), and slightly longer (315 pages versus 310 pages). A more notable difference is that 100 percent of the circulating books are in English, while only 94 percent of the noncirculating books are English language.

Differences in the average numbers of several access points are large enough to warrant testing. The large-scale test for comparing means described in Ott, Mendenhall, and Larson[13] was applied to the differences in selected means that exhibited the widest variation between circulating and noncirculating samples (see table 4). Because it was assumed from the raw data that more subjects and total access points might be associated with increased circulation, a one-tailed test was employed. The research hypothesis was that, for the variables of specific subjects, total subjects, and total access points, the means of the circulating subset of class *K* were greater than those of the noncirculating group. The research hypothesis was formulated as: $u_1 - u_2 > 0$. The null hypothesis was: $u_1 - u_2 = 0$. In order to reject the null hypothesis, with a 95 percent confidence interval, a value of z greater than 1.645 had to be obtained.

The values computed for two of the three tests surpassed the tabled figure, thus refuting the null hypothesis. The value of z was 1.8 for specific subjects, 1.68 for total subjects, and 0.78 for total access points. Thus it might seem that both the number of total subjects and the number of specific subjects play a role in higher circulation.

There is a danger in putting too much

TABLE 4

ANALYSIS OF CLASS K—CIRCULATING AND NONCIRCULATING BOOKS

(Underlined value is significant at .05 level)

Variable (average)	Circulating Books (76)	Noncirculating (156)	Total (232)	Correlation coefficients between variables and total circulation	
				Circulating subset (76)	Total (232)
Year of publication	1979	1977	1978	− .088	.068
Number of pages	315	310	312	− .053	− .015
English	100%	94%	96%	*	.104
Foreign		6%	4%		
Topical subject	1.84	1.78	1.80	.072	.050
Specific subject	.38	.18	.25	− .073	.067
Total subjects	2.22	1.96	2.05	.006	.088
Personal author	1.32	1.34	1.32	.058	.016
Corporate entries	.18	.17	.18	− .227	− .097
Series entries	.39	.50	.47	.108	− .026
Title entries	1.03	1.02	1.02	.194	.066
Total nonsubjects	2.92	3.03	2.98	.016	− .042
Total access points	5.14	4.99	5.03	.016	.028

*All circulating books were English language—no correlation possible.

emphasis on a single test such as this. While the means make it appear that a difference in the number of subject access points may account for higher circulation, the corresponding correlation coefficients cast doubt on this interpretation. The circulating subset of 76 books shows that the presence of more specific subjects is actually associated with *less* circulation (− .073), and total subjects have only a weak .006 positive correlation with total circulation. These differences indicate that while circulating books may on average have more subjects than noncirculating books, within the subset of circulating books those with more subjects do not necessarily circulate more often than those with fewer subjects. Factors such as chance, recommendations by faculty and peers, and qualitative judgments involved in choice of books, all of which may affect circulation, cannot be accounted for without more rigorous techniques of measurement. The quantitative factors that can be measured do not provide strong levels of association with circulation.

One constant within these statistics may provide a clue to their interpretation. The dichotomy in correlation coefficients when examined by type of subject and type of user appears again (see table 5). Undergraduates seem to use more books with more topical (general) subject headings, and academics use books with more of the specific subject headings. What may well be causing most of the difference in correlations is the type of materials preferred by the different user groups. Adding more access points to catalog records may have little association with circulation unless types of users are taken into account. This point deserves further study. If valid, one might (ideally) want to apply different norms of subject analysis depending on the clientele being served.

CONCLUSIONS

These figures should be considered in the context of other studies dealing with cataloging. Use studies have indicated that certain types of materials, such as foreign-language books, circulate less frequently than average.[14] A study of short entry catalogs done at Bath University showed that most patron needs could be met by brief entry records with no more than two access points.[15] The current study lends some weight to the idea that for certain types of cataloging the equivalent of a full MARC record is not cost efficient in terms of future circulation of the titles. If circulation is accepted as a valid measure of library performance, there is little to suggest that the time and expense of full local cataloging of older backlog items for which no MARC cataloging is

TABLE 5

CORRELATION COEFFICIENTS BY TYPE OF SUBJECT AND USER—CLASS K

(Underlined values have significant correlation at .05 level)

	Circulating Books (76)	All Books in Class K (232)
Topical Subjects:		
User Group		
Academic	− .120	− .062
Graduate	− .029	− .008
Undergraduate	.141	.082
Total*	.072	.050
Specific Subjects:		
User Group		
Academic	.171	.159
Graduate	.037	.067
Undergraduate	− .144	.010
Total*	− .073	.067
Total Subjects:		
User Group		
Academic	.029	.045
Graduate	.004	.036
Undergraduate	.010	.081
Total*	.006	.088

*Total figures include local nonacademic circulation

available, such as foreign-language materials, is justified.

Given the varied evidence gathered in this study there does not seem to be a simple, ready solution to the question of increasing subject accessibility. While it is true that most MARC records offer few subject access points, this research shows that merely increasing the number of LC subject headings may have little or no measurable effect on book retrieval and use. One intriguing possibility is that an incremental increase in subject headings, such as from one or two up to three or four, is not significantly associated with circulation, but that a very large increase up to perhaps ten or more—which would also allow access to subsets of information in monographs—may show an association with use figures. Carol Mandel has compiled a number of possible methods for such subject enhancement.[16] However, the data presented here, which is based on traditional cataloging records, does not offer statistical evidence that a larger number of subject headings is significantly associated with increased circulation.

It has been suggested that added entries are rarely consulted and might be dropped from catalog records under cer-

tain circumstances, especially if keyword title access were added.[17] In this study no type of added entry was positively associated with circulation at a significant level. When one considers the authority work that is often necessitated by series and corporate entries, consideration must be given to simplifying certain types of cataloging—older imprints, foreign-language materials, and various materials found in backlogs come immediately to mind.

More information, particularly protocol analysis and transaction log analysis, which can be more directly correlated with circulation, is necessary before making any major changes to the catalog record. The research reported here indicates that use patterns are complex and that catalog use varies from one subject area to the next and from one user group to the next. The catalog record of today may be a secondary factor in book selection for most users, compared to such other influences as browsing and booklists. The catalog of the future must be built on knowledge of actual user behavior, and should perhaps be keyed more to the institutional situation and less to traditional expectations. It is conceivable that a record with less de-

scription and fewer added entries will be combined with greater subject access of a type that will be simpler to use. As we move forward, creating new catalogs for a new generation of academic library users, every assumption concerning access should continually be tested. The knowledge gained will serve us well in the quest for ever better access to our collections.

REFERENCES

1. Margaret Ann Thomas Taylor, "The Effect of Bibliographic Accessibility upon Physical Accessibility in a Public Library Setting" (Ph.D. diss., Univ. of Michigan, 1982), p.19.
2. Pal V. Rao, "The Relationship between Card Catalog Access Points and the Recorded Use of Education Books in a University Library," *College & Research Libraries* 43:341–45 (July 1982).
3. William Carl Highfill, "The Relationship of Indexing Depth to Subject Catalog Retrieval Effectiveness" (Ph.D. diss., Univ. of Illinois, 1969).
4. Ibid., p.233.
5. Taro Yamane, *Statistics: An Introductory Analysis*, 2d ed. (New York: Harper, 1967), p.886.
6. *Basic Statistics for the Macintosh* (State College, Pa.: Time Zero Software, 1984).
7. Karen Markey, *Subject Searching in Library Catalogs: Before and After the Introduction of Online Catalogs* (Dublin, Ohio: OCLC, 1984), p.57–61.
8. Charles R. McClure, "Subject and Added Entries as Access to Information," *Journal of Academic Librarianship* 2:9–14 (Mar. 1976), p.10.
9. Edward T. O'Neill and Rao Aluri, "Library of Congress Subject Heading Patterns in OCLC Monographic Records," *Library Resources & Technical Services* 25:63–80 (Jan./Mar. 1981), p.70.
10. This is an altered and much simplified use of subject categories inspired by Stephen E. Wiberley, Jr., "Subject Access in the Humanities and the Precision of the Humanist's Vocabulary," *Library Quarterly* 53:420–33 (Oct. 1983).
11. Markey, p.108–9.
12. Renata Tagliacozzo, Lawrence Rosenberg, and Manfred Kochen," "Access and Recognition: From Users' Data to Catalogue Entries," *Journal of Documentation* 26:230–49 (Sept. 1970), p.234–40.
13. Lyman Ott, William Mendenhall, and Richard F. Larson, *Statistics: A Tool for the Social Sciences* (North Scituate, Mass.: Duxbury, 1978), p.258–64.
14. Alan Kent and others, *Use of Library Materials: The University of Pittsburgh Study* (New York: Dekker, 1979), p.44.
15. Alan Seal, "Experiments With Full and Short Entry Catalogues: A Study of Library Needs," *Library Resources & Technical Services* 27:144–55 (Apr./June 1983), p.149.
16. Carol A. Mandel, "Enriching the Library Catalog Record for Subject Access," *Library Resources & Technical Services* 29:5–15 (Jan./Mar. 1985), p.12.
17. Seal, p.147.