WILLIAM E. McGRATH

The Significance of Books Used According to a Classified Profile of Academic Departments

Classification Numbers of books were matched to a classified profile of the university teaching program descriptions. The profiles consisted of LC and DC numbers assigned to courses. It was found that book numbers which matched the course profiles were, (1) more likely to be charged out than not charged out, (2) after being removed from the shelves, more likely to be charged out than left on tables, (3) more likely to be taken off the shelves than left on. The differences between expected and actual proportions in these three situations are large enough to suggest that a precise, classified profile of the university program can be used successfully to select books and to predict circulation.

THERE HAS BEEN much interest recently in identifying factors which can be used to predict which books and how many will be most used and who will use them. If these factors were known and applied, librarians could be more confident that their book selections were appropriate and that they were building collections relevant to their college or university programs.

One recent study in this vein by G. Edward Evans concluded that books selected by librarians in the institutions sampled were more likely to circulate than those selected either by faculty or by an on-approval method.[1] Even if Evans' findings are true in general, we still do not know what it is about a book that enables a librarian (or any one else) to identify it as more "circulatable" than any other book.

This study stems from the premise that reliable conclusions about book usage can be reached by an examination of the characteristics of books themselves rather than of how they were acquired or who selected them. If it could be shown that a highly used book or group of books possessed characteristics that little-used books did not possess, then these characteristics could be built into a book selection policy. And as the characteristics change so could the policy. Fussler and Simon took this approach and showed that of several variables, immediate past use was the best single predictor of future use.[2] In this paper, the specific characteristic is the subject of books.

In previous work the author developed a technique for monitoring the collection by comparing book selection and circulation to a framework constructed from the university's catalog of courses —the classified course technique.[3] The list of classification numbers generated by the technique can be regarded as de-

Mr. McGrath is director libraries and professor of library science at the University of Southwestern Louisiana, Lafayette, Louisiana.

partmental subject profiles and hence as the university subject profile. Some doubts as to the effectiveness of the profile lingered, however, so a way to test it was sought. The question was simply: how accurate are the profiles; do they describe the departments and the university well enough to continue using them; are the profiles valid?

A plan was formulated to measure what happens to books with classification numbers which match those in the profile as compared to those for which the numbers did not match. Which books were used and which were not used?

The general collection of the University of Southwestern Louisiana library is open stack, all students and faculty may remove books from the shelves and use them in the library. Thus many books each day are left on tables.

The Hypotheses

The three conditions of use: (A) books charged out of the library; (B) books left on tables in the library; (C) books remaining on the shelves; are fundamental in formulating the hypotheses. Furthermore, we are interested in how the two contingencies, books whose classification numbers match or do not match those in the profile, affect the three conditions (A), (B), and (C). Thus the following basic hypotheses can be stated.

For books whose numbers match the profile, there is no significant difference between the proportion of books which are

I. (A) charged out vs. (B) and (C) not charged out,

II. (A) charged out vs. (B) left on tables vs. (C) left on shelves,

III. (A) charged out vs. (B) left on tables,

IV. (A) and (B) taken off the shelves vs. (C) left on the shelves,

V. (B) left on tables vs. (C) left on the shelves,

VI. (A) charged out vs. (C) left on the shelves,

VIII. (A) and (C) charged out or left on shelves vs. (B) left on tables.

If we found no significant differences among the conditions to be tested we would still not know for sure whether our profile was valid, and hence we would not know how much relevance the collection as developed over the years had to the university program, nor how much course content related to which books are used.

On the other hand, if significant differences were measured, the profile might be considered valid to the extent indicated by the difference between observed and expected proportions of actual use of the library. Furthermore, we would have some assurance that those parts of the collection in actual demand did reflect the university program, and that at least some of the demand stemmed directly from course content. The profile would thus be a valid tool for measuring these differences.

Method

Three independent data samples were collected by the author on three different occasions. The method for collecting the samples—counting the books falling within the LC or DC classification profile—is described in the author's paper on correlating books used in the library with those which go out.[4] The three samples were from (1) subject circulation of the University of Southwestern Louisiana books for nearly the entire academic year July 1969–May 1970; (2) one year's subject circulation from a study conducted by the author while at the South Dakota School of Mines and Technology in 1967/68; and (3) the study on correlation cited above.[5] Each of these samples required a count of the shelflist. Since no actual counts were available, estimates were made by measuring the shelflists at each institution, counting 100 volumes to the inch.

To match a book against the profile required no judgment, only a quick observation by the person doing the counting. Biases of match or nonmatch would be introduced by the original construction of the profile or by unknown factors causing a book to be removed from the shelves.

Since there are two contingencies of use, an appropriate design is the contingency figure, and an appropriate test is the chi-square test of independence. Only the first four hypotheses will be tested, since the last three are complements of the first four and the results can be readily deduced from them.

First Hypothesis

Books Charged (A) vs. Those Not Charged (B) and (C)

The counts in Tables 1 and 2 represent books charged or not charged during the two annual counts. "Not charged" could include books left on the shelves or on tables; i.e., no separate count for books left on tables was made. The "charged out" counts are what remained after all nonclassified materials (e.g., current periodicals) and permanent loans were removed. Charges did include books charged to the reserve reading room introducing a possible bias, since none of the samples considered the number of times a book on reserve was used.

Both figures then contain data for samples taken under essentially the same conditions but for two different institutions. Table 1 contains data from an eleven-month sample of the books charged from the library by undergraduates and graduate students at the University of Southwestern Louisiana from July 1969 to May 1970. Table 2 contains data from the South Dakota School of Mines and Technology.

TABLE 1

Books Charged Out (A) vs. Those Not Charged (B) and (C)
(U.S.L. Sample for 11 Months, 1969/70)

	Charged Out		Not Charged Out		
	Actual	Expected	Actual	Expected	Total
Match	53,333	42,905	76,550	86,977	129,883
Nonmatch	3,495	13,922	38,651	28,223	42,146
Total	56,828		115,201		172,029
Chi-square (χ^2)					15,448*
Roscoe's statistic (C_r)					.423

* Extremely significant at $\alpha_{.005}$; value of χ^2 needed for rejection with 1df is 7.88; $H_0{:}\phi = 0$ is thus rejected.

TABLE 2

Books Charged (A) vs. Those Not Charged (B) and (C) from the Library at the South Dakota School of Mines and Technology
(Sample for the Year, 1967/68)

	Charged Out		Not Charged Out		
	Actual	Expected	Actual	Expected	Total
Match	6,056	5,109	25,894	26,841	31,950
Nonmatch	1,640	2,587	14,539	13,592	16,179
Total	7,696		40,433		48,129
Chi-square (χ^2)					621.18*
Roscoe's statistic (C_r)					.159

* Highly significant at $\alpha_{.005}$; value of χ^2 needed for rejection with 1df is 7.88; $H_0{:}\phi = 0$ is rejected.

In both libraries, the chi-square statistic indicates that for the yearlong samples, there is a statistically significant difference between the proportion of books charged and not charged. The chi-square statistic is highly significant in Table 2 and extremely significant in Table 1. In both samples, the hypotheses are overwhelmingly rejected. For example, in Table 1, at the University of Southwestern Louisiana, we expect the proportion of matching books charged out to be 42,905/129,883, or about 33 percent, but in actuality a larger proportion 53,222/129,883, or about 43 percent was charged. Conversely, we expect the proportion of nonmatching books charged out to be 13,922/42,146 or again 33 percent, but instead 3,495/42,146, or about 8.3 percent were charged, a much smaller proportion.

Similarly, in Table 2 for the South Dakota School of Mines and Technology, we expect a proportion of matching books charged out to be 5,109/31,950, or about 16 percent, whereas in actuality a proportion of 6,056/31,950 or about 19 percent was charged. Again, the proportion of nonmatching books charged is smaller than expected.

In both libraries, according to the significant value of χ^2 we can expect books with numbers that match the profile to have a greater chance of being charged out. Another way of looking at the actual differences is through some statistic which measures the degree of effect, or contingency. In this paper, we have used Roscoe's statistic,[6]

$$C_r = \sqrt{\left(\frac{k}{k-1}\right)\left(\frac{\chi^2}{n+\chi^2}\right)}$$

where if χ^2 were 0, or small, C_r would be small, and if there were a perfect relationship between the contingencies—i.e., if all matching books circulated and all nonmatching books did not circulate, C_r would be large or near 1.0. Interpretation of C_r is subjective. In Table 1, C_r is substantial; in Table 2, the effect is much less. χ^2 statistic used by itself is misleading because significance is almost guaranteed with the large N's used here.

Second Hypothesis

Books Charged Out (A) or Left on Tables (B) or Left on Shelves (C)

Librarians and faculty have long doubted that charge statistics reflect the true use of libraries. Charge statistics usually do not include counts of books in the library, which may reveal a different use pattern. The two samples treated in Hypothesis 1 did not break down the data to include a count of in-library use. An in-library count, however, was included as part of the author's study on correlation.[7] A one-month count of all books left on tables, chairs, in restrooms, and other locations was conducted. Books which were removed from the shelves by users were counted at the time of reshelving by student aids. The in-library count and the charge count were made during the same period. The three-way count appears in Table 3.

This hypothesis states that we expect no differences among the three proportions for matching books; those charged out, those left on tables, and those left on shelves. When the three conditions are evaluated together, the chi-square value of 443.13 indicates that there is a statistically significant difference among the three proportions. This statistic, however, reflects the overall difference. It does not tell us whether the individual differences are significantly larger or significantly smaller, and the effects of the contingencies are cancelled out as shown by the small C_r statistic. Dual comparisons must therefore be made, where the data for the three conditions are partitioned or combined so that two conditions are compared, as under Hypothesis 1.

THIRD HYPOTHESIS

Books Taken Out (A) vs. Those Left on Tables (B)

Once a book is removed from the shelves, it can be placed in one of two categories. The book is (1) charged and taken out, or (2) the book is left on a table and/or returned to the shelves. Since we are seeking to validate the profile, we are definitely interested in knowing whether the profile has any bearing on how a book is used—even after it has been taken off the shelf. Here the hypothesis states that there is no difference between the proportion of books with matching numbers that are charged out or that are left on tables in the library.

As with the first two hypotheses, a statistically significant difference between the proportion of books charged out and those left on tables is found. With a chi-square value of 423.6, Hypothesis 3 is substantially rejected. We expect 6,900 matching books to have been charged out but instead 7,385 were charged; and whereas 2,052 nonmatching books should have been charged, only 1,568 were charged. The significance, of course, is in the difference between the actual and the expected proportions. In general, if a book's classification number matched a number in the profile, there was a greater probability that it would be charged out after having been removed from the shelves. If its number did not match, there was a greater probability that it would be left on the tables. Roscoe's statistic shows that the profile has a moderate effect on the probabilities.

FOURTH HYPOTHESIS

Books Taken Off the Shelves (A) and (B) vs. Those Remaining On (C)

This hypothesis is the complement of Hypothesis 3, where the total for books taken off the shelves (Table 5) equals

TABLE 3

BOOKS CHARGED OUT (A), LEFT ON TABLES (B), AND LEFT ON SHELVES (C) (One-Month Sample)

	Charged Out		Left on Tables		Left on Shelves		
	Actual	Expected	Actual	Expected	Actual	Expected	Total
Match	7,385	6,900	2,989	3,473	119,509	119,720	129,883
Nonmatch	1,568	2,052	1,518	1,033	39,060	38,849	42,146
Total	8,953		4,507		158,569		172,029
Chi-square (χ^2)							443.136*
Roscoe's statistic (C_r)							.072

* Significant at $\alpha_{.005}$; value of χ^2 needed for rejection with 2df is 10.6; $H_0:\phi = 0$ is thus rejected. The χ^2 value of 443.136 equals the sum of χ^2 values 423.623 and 19.5128 in Tables 3 and 4.

TABLE 4

BOOKS CHARGED FROM THE LIBRARY (A) VS. THOSE LEFT ON TABLES (B) (One-Month Sample)

	Charged Out		Left on Tables		
	Actual	Expected	Actual	Expected	Total
Match	7,385	6,900	2,989	3,473	10,374
Nonmatch	1,568	2,052	1,518	1,033	3,086
Total	8,953		4,507		13,460
Chi-square (χ^2)					423.623*
Roscoe's statistic (C_r)					.218

*Highly significant at $\alpha_{.005}$; value needed for rejection with 1df is 7.88; $H_0:\phi = 0$ is rejected.

the sum of books charged out plus those left on tables as shown in Table 4. It is necessary to combine the two in this manner to account for the degrees of freedom. The chi-square value of 19.5 is again significant, but as can be seen from the expected number of matching books removed from the shelves, and the small value of C_r the effect is not so readily apparent, thus pointing up the need for distinguishing between books actually charged and those left on tables.

Implications

With all four hypotheses substantially rejected, the conclusions are tempting: the profile does describe, within the limits of probability derived from the differences in actual and expected proportions, the books used; this usage is clearly related to the subjects embraced by the university's academic departments; the profile is therefore a valid predictor of usage.

These conclusions cannot be drawn unequivocally on the basis of only three samples. That is, for greater confidence, additional samples should be drawn from other libraries under carefully controlled conditions.

The pattern of differences revealed by the three combinations of conditions discussed in hypotheses 1, 2, and 4 suggest the following. If a student or faculty member (we don't know which) removes a book from the shelves he is more likely to charge it out if its class number happens to match the profile (hypotheses 3 and 4). This is not to say that he examines the call number to ascertain the subject of the book. As a matter of speculation, the call number may be nothing more to the user than a location device. If a book is removed from the shelf and left on a table, there is a greater probability that its class number does not match the profile. That is, a person must remove the book from the shelf, and examine it before he knows whether he wants it. This is indicated by the fact that he has taken the time to bring it to a table. The implication here is that most books left on tables may be those of which the users are unsure. So-called "in library" use, at least in an open stack library where users have a choice of taking books out or using them in the library may not constitute real use, at least in some subject areas. Such "use" may actually be "to see whether I want to use the book," and therefore should not be equated with out-of-library use. On the other hand, to draw a severe distinction between the two types of use may be stretching the point. Even though we can now describe to a certain extent which subjects students will take out, we do not know for sure why they take them.

We do not suggest that matching numbers "cause" a book to be taken out nor do we suggest that the matching number is the only, or the best, indicator. We do suggest, however, since the differences between expected and ac-

TABLE 5

Books Taken Off the Shelves (A) and (B) vs. Those Remaining On (C)

	Books Taken Off		Books Remaining On		
	Actual	Expected	Actual	Expected	Total
Match	10,374	10,162	119,509	119,720	129,883
Nonmatch	3,086	3,298	39,060	38,848	42,146
Total	13,460		158,569		172,029
Chi-square (χ^2)					19.5128*
Roscoe's statistic (C_r)					.012

*Significant at $a_{.005}$; value needed for rejection with 1df is 7.88; $H_0:\phi = 0$ is rejected.

tual difference are statistically significant, even though the differences are not large, that the classified profile is a significant indicator in the two libraries studied and can be used to predict which subjects will be most in demand. The success of a profile of course depends to no small degree on its accuracy.

The findings suggest if a library owns more books for which the classification numbers matched those of the teaching program, then a greater proportion of the library's titles would be used. As it happens, 75 percent of the University of Southwestern Louisiana library's collection matches its profile—not bad it would seem for a nonsystematic, informal history of random selection and collection building. But still not good enough if we are to believe our statistical results.

The implication is clear: if books in the collection are more likely to be used when their numbers match the profile, then why not in the beginning add books whose numbers match? As for multiple copies, many books charged out are multiple charges of a smaller number of titles. These charges do not invalidate the profile; rather they support the profile.

Why not use the classification profile then as a selection aid if not an outright criterion? The University of Southwestern Louisiana library, for example, matches its profile against the monthly or annual issues of the *American Book Publishing Record* as a selection aid for certain academic departments, and plans to use it with the MARC tapes, as the Oklahoma State library is doing with its profile of state agencies.[8]

Outspoken criticism of blanket-order plans in libraries with limited budgets suggest that such plans need reexamination.[9] If a library can show that a carefully constructed, precise profile of its program, used in combination with other delineators such as class level, publisher, and language, describes or embraces the subjects of books actually used, then such profiles contribute in efforts to reduce the number of undesired books received through existing plans.

A few scholars and some must-build-a-great-library librarians who strive to acquire a copy of every book would throw up their hands in horror at the suggestion that we ignore many fine books not in demand and purchase only books presently in demand. But, in a medium-sized university with a limited book budget, can we afford to slight legitimate demands of the curriculum while catering to the esoteric demands of the scholar? Certainly not if we agree with Grant who has found that, "with very few exceptions, students are apparently checking out only books that are curriculum oriented in the most narrow sense. . . ."[10]

Further Research

The samples discussed in this paper were for overall circulation. It would be useful to know (1) whether the same results hold true for the individual departments, such as those listed in the study on correlation cited above, (2) what proportion of books which both match a classified profile and circulate, are accounted for by books published within a given imprint year, (3) to what extent charges to reserve and their circulation account for books matching the profile, and (4) whether the same results hold true for other libraries.

References

1. G. E. Evans, "Book Selection and Book Collection Usage in Academic Libraries," *Library Quarterly* 40:297–308 (1970).
2. H. E. Fussler and J. L. Simon, *Patterns in the Use of Books in Large Research Libraries* (Chicago: University of Chicago, 1969).
3. W. E. McGrath and N. Durand, "Classify-

ing Courses in the University Catalog," *CRL* 30:533–39 (1969).
4. W. E. McGrath, "Correlating the Subjects of Books Taken out of and Books Used Within an Open-Stack Library," *CRL* 32: 280–85 (1971).
5. Ibid.
6. J. T. Roscoe, *Fundamental Research Statistics for the Behavioral Sciences* (New York: Holt, Rinehart and Winston, Inc., 1969), p.202.
7. McGrath, "Correlating . . ."
8. K. J. Bierman, "An Operating MARC-based SDI System: Some Preliminary Services and User Reactions," *Proceedings of the American Society for Information Science* 7:87–89 (1970).
9. R. Rouse, "Automation Stops Here: A Case for Man-made Book Collections," *CRL* 31:147–54 (1970).
10. R. S. Grant, "Predicting the Need for Multiple Copies of Books," *Journal of Library Automation* 4:64–71 (1971).