

## Sampling and Short-Period Usage in the Purdue Library

*Several possible methods of sampling the social science monograph titles in the general library of Purdue University were considered, and a "good" method was used to obtain estimates of their usage in the library and at home during the period July 1–August 4, 1964. The term relative usage was defined and used to study the effect of: (1) language, (2) country of publication, (3) year of publication, and (4) year of accession of a monograph title. An attempt was made to fit a regression model for titles in English by quantifying the last three independent variables with relative usage as the dependent variable. Functions based on the above variables have been developed to identify monograph titles for storage.*

*A questionnaire was employed to study the usage of library facilities and to gather opinions of library patrons. Purpose of visiting the library, reason for checkout of library material, reason for preferring library or home for the use of library material, etc., were analyzed on the basis of the replies received.*

THE USUAL APPROACH to studying usage in a library is to start with a representative sample of titles, to record usage of these titles from their book cards (assuming such records are available) and to analyze these data. The study by Fussler and Simon<sup>1</sup> is an excellent example. Trueswell<sup>2</sup> has studied the "last circulation date" for books by saving the

book cards for each day's circulation and has derived some useful rules for thinning the stacks. A question remaining unanswered is what a librarian who does not have book cards does to study usage of books? This paper presents an approach to studying usage when there are no book cards. Also, a librarian who has book cards may prefer to use this approach in place of the usual approach because of its simplicity and convenience.

This paper also presents some results of a questionnaire survey carried out to study the usage of the Purdue library facilities and to gather opinions of library patrons.

### THE APPROACH

Three independent samples of monograph titles from the total collection

<sup>1</sup>Herman H. Fussler and J. L. Simon, *Patterns in the Use of Books in Large Research Libraries* (Chicago: University of Chicago Library, 1961).

<sup>2</sup>Richard W. Trueswell, "A Quantitative Measure of User Circulation Requirements and Its Possible Effect on Stack Thinning and Multiple Copy Determination," *American Documentation*, XVI (January 1965).

---

Mr. Jain is in the School of Industrial Engineering at Purdue University. This paper was read at the 1965 annual meeting of the American Society for Engineering Education in Chicago.

(S), home-used material (H), and in-library used material (I) respectively are obtained. These samples are divided into a certain number of groups on the basis of the following characteristics of the title: (1) language, (2) country of publication, (3) year of publication, and (4) year of accession. The frequencies of titles in these groups are computed for each of the three samples. Let  $S_i$  = number of titles in sample S which belong to the  $i$ th group;  $H_i$  = number of titles in sample H which belong to the  $i$ th group;  $I_i$  = number of titles in sample I which belong to the  $i$ th group. Then, relative usage  $R$  is defined as follows:

$$R_i = \% \text{ relative usage of } i\text{th group} = \frac{H_i + I_i}{S_i} (100).$$

The magnitude of the indices  $R_i$  depends on the relative sizes of the samples S, H, and I. For example, if we take a very small sample of total collection (S) but take all titles used at home (H) and in-library (I) during a long period of time (say five years) as the other two samples, then indices  $R_i$  will be large in magnitude. But, in any case,  $R_i/R_j$  is independent of all possible variations in samples S, H, and I, for all  $i$  and  $j$ , as long as

$$\frac{\sum H_i}{\sum I_i}$$

is fixed. Because of this, it is very useful to compare these indices  $R_i$  among themselves and the groups having the lowest  $R_i$  are the ones which are the least important from the point of view of usage.

We will first discuss the method of sampling and then discuss some results based on the above indices  $R_i$ .

#### SAMPLING TOTAL COLLECTION, HOME USE, AND IN-LIBRARY USE

This work was done during the summer session (June 15–August 7) 1964. Since time was short, it was decided to confine the scope of this study to mono-

graph titles in Dewey Decimal Classification (DDC) 330-379. A "title," including all copies of all volumes of all editions on the third floor of the Purdue general library, was considered as the best sampling unit. The sampling schemes used to obtain the three samples S, H, and I are discussed below.

*Total Collection (S).* The following sampling frames were considered: (1) books on shelves on the floor; (2) author catalog; (3) subject and title catalog; and (4) shelf list. The first frame was rejected because all books can not be found on shelves at any point of time (our study has revealed that only 65 per cent of titles were on shelves) and because it was not handy. The second frame was considered good since it was easy to get a representative sample of the titles (assign weights inversely proportional to the number of authors, *i.e.* number of cards in author catalog). For this study, however, it would have been very wasteful because there was interest in DDC 330-379 only. The third frame sounded good. But it turned out that there were many titles and subtitles for each subject and it would have been extremely difficult to get a representative sample of DDC 330-379.

The last frame, *viz.* shelflist, was considered the best. The only drawback of the shelflist frame was that the number of cards for a title was not necessarily one (in some cases there were more than five cards for a title). This drawback was remedied by ignoring all cards except the first one for each title. There were fifty-one drawers of shelflist cards for DDC 330-379. Since the time was short, it was decided to obtain a 5 per cent (or 1 in 20) sample of titles.

The next question was what sampling method to use, systematic or random? Systematic sampling was chosen because Fussler and Simon tested a subject area for cyclical effects in systematic sampling and found none and because ran-

TABLE 1

LIBRARY'S HOLDINGS AND RELATIVE USAGE BY YEAR OF PUBLICATION

YEAR OF PUBLICATION	NUMBER S IN DDC			PER CENT S IN DDC			PER CENT R IN DDC		
	330-339	340-369	370-379	330-339	340-369	370-379	330-339	340-369	370-379
Pre-1904 . . . . .	14	12	13	3.2	3.4	3.0	43	42	231
1904-1913 . . . . .	28	18	27	6.3	5.1	6.2	46	22	274
1914-1923 . . . . .	36	17	41	8.1	4.8	9.4	39	24	85
1924-1933 . . . . .	45	30	42	10.1	8.4	9.6	80	43	148
1934-1943 . . . . .	73	62	58	16.5	17.4	13.2	86	44	238
1944-1953 . . . . .	105	82	81	23.7	23.1	18.5	118	77	347
1954-1964 . . . . .	139	133	169	31.4	37.5	38.7	165	135	598
Not Available . . . . .	3	1	6	0.7	0.3	1.4	67	0	67
Total . . . . .	443	355	437	100.0	100.0	100.0	110	83	374

dom sampling would have been very expensive. To get a 5 per cent sample nineteen titles had to be skipped between every two titles in the sample. It would have been time-consuming to skip nineteen titles, and there was a good possibility of error in count, since cards other than the first were to be ignored. After considering various possibilities the method used was the following: each of the fifty-one drawers was divided into four equal parts (*i.e.* four equal lengths of cards), and one out of the four parts was selected at random. In the selected part, a random start was made by picking up one of the first five cards at random including every fifth title (systematically).

*Home Use (H).* There are no book cards in Purdue library. A checkout slip is filled in at the time of checkout of a title, and it is discharged after the title is returned to the library. Checkout slips for titles in DDC 330-379 returned to the library during July 1-August 4, 1964 were saved, and these titles constituted sample H. This type of sample is different from a sample consisting of titles checked out during a certain period. Both of these are subject to a bias because of variations in loan periods for different classes of users. The former was chosen because it was easy to handle discharged slips.

*In-Library Use (I).* Library patrons pick up titles from the open shelves of the library and leave them on tables after their use. Though not very realistic, the simplifying assumption was made that all titles left on tables had been used (*i.e.* the fact was ignored that some of the titles left on tables might have been "rejected" instead of "used" by the library patrons). All titles in DDC 330-379 found on tables early in the morning and late in the evening during July 2-August 4, 1964 constituted sample I. This sample accounted for about 67 per cent of the total material (DDC 330-379) reshelved during the survey period.

#### RESULTS BASED ON SHORT-PERIOD USAGE

The over-all relative usages of monograph titles in DDC 330-339, 340-369, and 370-379 were 110, 83, and 374 respectively. Education (*i.e.* DDC 370-379) titles had three to four times as much usage as titles in DDC 330-339 and DDC 340-369. This was mainly because many high school teachers took courses in education at Purdue during the summer session. The pattern is expected to be quite different during the regular semesters.

*Use and Age of Title.* Table 1 presents the Purdue library's holdings and relative usage by year of publication. It

TABLE 2  
LIBRARY'S HOLDINGS AND RELATIVE USAGE BY YEAR OF ACCESSION

YEAR OF ACCESSION	NUMBER S IN DDC			PER CENT S IN DDC			PER CENT R IN DDC		
	330-339	340-369	370-379	330-339	340-369	370-379	330-339	340-369	370-379
Pre-1924 . . . . .	52	29	45	11.7	8.2	10.3	23	14	184
1924-1933 . . . . .	33	25	41	7.4	7.0	9.4	94	24	83
1934-1943 . . . . .	75	50	69	16.9	14.1	15.8	77	40	203
1944-1953 . . . . .	80	69	73	18.1	19.4	16.7	110	61	203
1954-1964 . . . . .	197	177	197	44.5	49.9	45.1	148	126	621
Not Available . . . . .	6	5	12	1.4	1.4	2.7	117	0	42
Total . . . . .	443	355	437	100.0	100.0	100.0	110	83	374

will be seen that, generally speaking, relative usage decreased monotonically with age (*i.e.* the number of years since the date of publication) for DDC 330-339 and DDC 340-369. But for DDC 370-379 the pattern was quite different: titles published before 1904, during 1904-1913, and 1934-1943 had about the same relative usage which was higher than the relative usage for titles published during 1914-1933. Table 2 shows that the pattern by year of accession was similar to that by year of publication.

*Use of Foreign Books.* From Table 3 it is clear that the relative usage was higher for titles published in the USA and England than for those published in France, Germany, and other countries. It may be pointed out that in the Purdue library about 85 per cent of titles in DDC 330-339 and DDC 340-369 and

94 per cent of titles in DDC 370-379 were published in the USA.

Table 4 shows that the relative usage was the highest for titles in English for each of the three groups. The next important language from the point of view of usage was French. It is interesting to note that all used titles (*i.e.* samples H and I) in DDC 340-369 were in English.

It must be mentioned that since 94 per cent or more of the titles in the Purdue library in DDC 330-379 are in English, it is necessary to use larger sample sizes before any definite conclusions can be drawn regarding relative usage of titles in languages other than English.

*Usage by Class of User.* The Table 5 shows the home usage of monograph titles in DDC 330-379 by class of user. It is clear that graduate students used

TABLE 3  
LIBRARY'S HOLDINGS AND RELATIVE USAGE BY COUNTRY OF PUBLICATION

COUNTRY	NUMBER S IN DDC			PER CENT S IN DDC			PER CENT R IN DDC		
	330-339	340-369	370-379	330-339	340-369	370-379	330-339	340-369	370-379
USA . . . . .	378	306	410	85.3	86.2	93.8	111	91	387
England . . . . .	23	26	6	5.2	7.3	1.4	222	46	667
France . . . . .	8	4	4	1.8	1.1	0.9	50	25	100
Germany . . . . .	5	8	4	1.1	2.3	0.9	20	0	0
Other . . . . .	29	11	13	6.6	3.1	3.0	34	45	15
Total . . . . .	443	355	437	100.0	100.0	100.0	110	83	374



TABLE 4  
LIBRARY'S HOLDINGS AND RELATIVE USAGE BY LANGUAGE OF TITLE

LANGUAGE	NUMBER S IN DDC			PER CENT S IN DDC			PER CENT R IN DDC		
	330-339	340-369	370-379	330-339	340-369	370-379	330-339	340-369	370-379
English . . . . .	417	338	427	94.1	95.2	97.7	116	87	382
French . . . . .	9	4	5	2.0	1.1	1.2	44	0	60
German . . . . .	4	10	4	0.9	2.8	0.9	25	0	0
Other . . . . .	13	3	1	3.0	0.9	0.2	0	0	0
Total . . . . .	443	355	437	100.0	100.0	100.0	110	83	374

the library three times as much as undergraduates and four times as much as faculty. This is in close agreement with the findings of Snyder.<sup>3</sup>

Where Were the Titles? Every librarian would like to know where his books are: how many are checked out, how many are on shelves, etc. We carried out an inventory of titles in sample S for Education (DDC 370-379) on July 22-23 and found the following (Table 6). In Table 6 "not traceable" means not accounted for by any of the previous categories. It was thought that some of these might be back<sup>4</sup> on shelves after the summer session and accordingly two more rounds were made on August 14 and 17 respectively. During these rounds twenty-two out of eighty-one "not traceable" titles were found, and the remaining were considered either missing or misshelved. The estimate of the availability on shelves (65 per cent) is quite close to that of Trueswell.

REGRESSION MODELS AND STORAGE FUNCTIONS

Originally, the plan was to quantify the four factors (1) language, (2) country, (3) year of publication, and (4) year of accession. Unfortunately, this could not be done mainly because (1)

94 per cent or more titles were in English, (2) 85 per cent or more titles were published in the USA, and consequently (3) there were very few, if any, titles in the samples belonging to some of the thirty classifications based on the four factors considered above (especially with "Not English"). The best that could be done was to fit a regression model for English-language titles. The following model for the English titles in DDC 330-339 was obtained:

$$R = 181 + 1619x_1 - 112x_2 - 62x_3 - 107x_4 - 81x_5 - 134x_6 - 1538x_1x_3 -$$

TABLE 5

USER		NUMBER OF TITLES USED	
Class	Number	Total	Per 100 Persons
Undergraduate . . . . .	2,444	216	8.8
Graduate . . . . .	3,497	875	25.0
Faculty . . . . .	1,023	66	6.5
Other . . . . .	3,821	368	9.6
Total . . . . .	10,785	1,525	14.1

TABLE 6

WHERE?	SAMPLE S	
	Number	Per Cent
On shelves . . . . .	286	65.4
Checked out . . . . .	20	4.6
On reserve or reference . . . . .	26	6.0
On indefinite loan . . . . .	20	4.6
Known missing in 1960 . . . . .	4	0.9
Not Traceable . . . . .	81	18.5
Total . . . . .	437	100.0

<sup>3</sup>Helen I. Snyder, "Toward an Optimal Library System for Pennsylvania State University." Paper presented at American Society for Engineering Education annual meeting, June 1965.

<sup>4</sup>Could be due to in-library use, in transit, records under process, etc. during July 22-23.

TABLE 7

LIBRARY'S HOLDINGS AND USAGE (H, I, R) BY LANGUAGE, COUNTRY, YEAR OF PUBLICATION AND YEAR OF ACCESSION

DDC	LANGUAGE	COUNTRY	YEAR OF PUBLICATION	YEAR OF ACCESSION	NUMBER IN SAMPLE			PER CENT	
					S	I	H	S	R
330-339	English	USA	1954-64	Any	110	45	154	24.8	181
			1944-53	Any	90	40	67	20.3	119
			1924-43	Any	103	25	51	23.3	74
			Up to 1923	After 1943	8	7	1	1.8	100
			Up to 1923	Up to 1943	64	17	13	14.4	47
	Not English	England	After 1923	Any	17	25	22	3.8	276
			Up to 1923	Any	6	2	2	1.4	67
			Any	Any	19	3	8	4.3	58
			Any	Any	26	4	1	5.9	19
			1954-64	Any	111	61	107	31.3	151
340-369	English	USA	1944-53	Any	70	33	27	19.7	86
			1924-43	Any	88	28	9	24.8	42
			Up to 1923	After 1943	2	4	1	0.6	250
			Up to 1923	Up to 1943	33	2	5	9.3	21
			Any	Any	26	6	6	7.2	46
	Not English	England	1954-64	Any	3	2	4	0.9	200
			Up to 1953	Any	5	0	0	1.4	0
			Any	Any	17	0	0	4.8	0
			1954-64	Any	157	363	625	35.9	629
			1944-53	Any	75	94	180	17.2	365
370-379	English	USA	1924-43	Any	95	86	106	21.7	202
			Up to 1923	After 1943	20	4	12	4.6	80
			Up to 1923	Up to 1943	62	31	87	14.2	190
			Any	Any	6	9	31	1.4	667
			Any	Any	12	0	3	2.8	25
	Not English	England	Any	Any	10	0	3	2.2	30

$1531x_1x_4 - 1669x_1x_5 - 1591x_1x_6 - 7x_2x_3$   
 $+ 238x_2x_4 + 12x_2x_5 + 65x_2x_6,$   
 where R = relative use; variables  $x_1, x_2$  quantify country and  $x_3, x_4, x_5, x_6$  quantify age as follows.

Country	$x_1$	$x_2$
USA . . . . .	0	0
England . . . . .	1	0
France, Germany and Other . . . . .	0	1

Year of Publication	Year of Accession	$x_3$	$x_4$	$x_5$	$x_6$
1954-1964 . . . . .	Any . . . . .	0	0	0	0
1944-1953 . . . . .	Any . . . . .	1	0	0	0
1924-1943 . . . . .	Any . . . . .	0	1	0	0
Up to 1923 . . . . .	After 1943 . . . . .	0	0	1	0
Up to 1923 . . . . .	Up to 1943 . . . . .	0	0	0	1

Similar regressions have been worked out for DDC 340-369 and DDC 370-379.

It must be pointed out that since the number of titles in samples S, I, and H is rather small for quite a few of the fifteen cells used in the above regression model, the estimates of the corresponding regression coefficients are subject to large sampling errors. The magnitude of these sampling errors, of course, can be reduced by taking larger samples.

The library's holdings and relative usage by language, country, year of publication, and year of accession are presented in Table 7. The classification based on the above four factors is as detailed as possible subject to the small samples S, I, and H. This table is both interesting and useful. Relative usages in this table are the estimates of the regression coefficients.

Based on Table 7 (or the multiple linear regression) the functions given in

Table 8 have been constructed to identify Economics (DDC 330-339) monograph titles for possible storage. Similar functions have been constructed for DDC 340-369 and DDC 370-379. It is encouraging to note that the results of storage functions for DDC 340-379 are quite similar to those in Table 8 for DDC 330-339. These decision rules are relatively simple to apply and easy to communicate to the patrons. However, the evaluation of these functions has to be made by the administrators of the library with respect to its over-all goals.

USAGE OF LIBRARY FACILITIES

The Johns Hopkins University<sup>5</sup> has carried out a survey of library usage to construct a picture of the activities which

<sup>5</sup> The Johns Hopkins University, "Progress Report on an Operations Research and Systems Engineering Study of a University Library," April 1963.

make up a library day. A questionnaire survey was also conducted in a part of the Purdue general library (not restricted to social sciences only) to study the patrons' purposes of visit and preferences, etc. This survey was in operation for a few hours on each day during July 21-August 7, 1964 and 212 completed questionnaires were received.

Table 9 summarizes the replies regarding the purpose of visit to the library. In (ii) a patron has been counted more than once if he reported more than one purpose. "Use of own material," "use of library material," and "check out for home use" were mentioned by 60 per cent, 54 per cent, and 20 per cent of the persons respectively. About 8 per cent reported "other" (or personal business). According to the Johns Hopkins study about 10 per cent of the patrons con-

TABLE 8  
STORAGE FUNCTIONS FOR ECONOMICS (DDC 330-339) MONOGRAPH TITLES

Storage Function	Per Cent Titles Stored	Upper Cutting Point	Per Cent Use Generated by This Group
1. Use as a function of publication date (assuming monotonicity)	(i) 10	1915	4.3
	(ii) 20	1927	8.6
	(iii) 30	1935	15.8
2. Use as a function of accession date (assuming monotonicity)	(i) 10	1921	2.1
	(ii) 20	1934	9.4
	(iii) 30	1940	16.2
3. Use as a function of publication date excluding post 1943 accessions	(i) 10	1917	4.1
	(ii) 20	1929	8.2
	(iii) 30	1936	15.6
4. Use as a function of publication date, accession date, language and country (multiple linear regression)	(i) 6	All not in English	1.0
	(ii) 20	All in (i); English titles published in USA before 1924 and accessioned before 1944	7.2
	(iii) 26	All in (ii); English titles published in England before 1924; English titles published in other countries	10.2

TABLE 9  
PURPOSE OF VISIT TO THE LIBRARY

PART	PURPOSE	PERSONS	
		Number	Per Cent
(i)	Use of own material only . . . . .	69	33
	Use of library material only . . . . .	44	21
	Check-out for home use only . . . . .	7	3
	Other only . . . . .	17	8
	Use of library and own material . . . . .	40	19
	Use of library material and check-out . . . . .	17	8
	Use of own material and check-out . . . . .	5	2
	Use of library and own material and check-out . . . . .	13	6
	Total . . . . .	212	100
	(ii)	Use of own material . . . . .	127
Use of library material . . . . .		114	54
Check-out for home use . . . . .		42	20
Other . . . . .		17	8

ducted personal business, and about 50 per cent used their own material.

When asked "If checking out items *now*, did you intend when you came, to borrow them or did you get interested in them as a result of browsing?", "came to borrow," "result of browsing," and "both" were mentioned by 59 per cent, 34 per cent, and 7 per cent respectively of those who replied to this question.

About 46 per cent of library patrons "preferred" and 29 per cent "did not prefer" to use library material in the library rather than checking it out for home use. The reasons for those preferring the library were: better study atmosphere (46 per cent), to avoid mislaying of material (17 per cent), easier to refer to other sources (12 per cent), save the trouble of carrying it home (8 per cent), etc. Similarly, the reasons for those not

TABLE 10

REASON	PERSONS	
	Number	Per Cent
Own interest . . . . .	69	45
Course assignments . . . . .	52	34
Needed for term paper . . . . .	51	34
Study for course exam . . . . .	28	18
Other . . . . .	16	11
Total* . . . . .	152	100

\* The actual sum of "Persons" exceeds total because of multiple reasons.

preferring the library were: more comfortable at home (40 per cent), need for longer period (29 per cent), can use at leisure (16 per cent), use in conjunction with own material or typewriter (7 per cent), etc.

The library patrons who used library material during "this visit," were asked the reason for use of the library material in the library. The replies were as follows (Table 10).

On an average a library patron spent 2.2 hours in the library and used three titles during one visit. Thus, he spent approximately 0.7 hours per title used in the library. This may be compared with the usage of 7.7 hours per title at home obtained by a survey of material used at home during the same period.

#### CONCLUSION

This investigation was mainly exploratory and will be used as a guide for further and more complete studies of usage in the Purdue library. As mentioned earlier, patterns of use in fall and spring may be quite different from those in summer. It is therefore advisable to analyze data on use throughout the year. Also, it is necessary to take a larger sample to reduce the magnitude of sampling error. ■ ■