DOCUMENT RESUME

ED 110 025 IR 002 294

AUTHOR Rouse, William B., Ed.
TITLE Quantitative Approaches to the Management of Information/Document Retrieval at the University of Illinois.
INSTITUTION Illinois Univ., Urbana. Dept. of Mechanical and Industrial Engineering.; Illinois Univ., Urbana. Graduate School of Library Science.
PUB DATE Jun 75
NOTE 48p.

EDRS PRICE MF-$0.76 HC-$1.95 PLUS POSTAGE
DESCRIPTORS Data Collection; *Documentation; Geographic Distribution; Information Dissemination; *Information Retrieval; *Information Systems; Interlibrary Loans; Library Reference Services; *Library Research; Library Services; Library Surveys; Models; Research Projects; *University Libraries
IDENTIFIERS *University of Illinois

ABSTRACT
Three papers based on projects produced in a course entitled Operations Research and Library Management, jointly sponsored by the Department of Mechanical and Industrial Engineering and the Graduate School of Library Science are reported and explained. Topics covered include an assessment of faculty interest in an information retrieval service; modeling closed-stacks document retrieval, and the effect of geographic dispersion of the collection on document retrieval time. (SK)

QUANTITATIVE APPROACHES TO THE MANAGEMENT OF INFORMATION/DOCUMENT RETRIEVAL AT THE UNIVERSITY OF ILLINOIS

Edited By

William B. Rouse

Department of Mechanical and Industrial Engineering

June 1975

Department of Mechanical and Industrial Engineering

and

Graduate School of Library Science

University of Illinois at Urbana-Champaign

FOREWARD

These three papers are based on projects done in conjunction with a course entitled Operations Research and Library Management which was jointly sponsored by the Department of Mechanical and Industrial Engineering and the Graduate School of Library Science. One third of the class time was devoted to discussion of these projects while the remaining two thirds was devoted to lectures on probability and statistics, queueing systems, mathematical programming, and advanced applications.

The text used was Swanson and Bookstein (Editors), Operation Research: Implications for Libraries, University of Chicago Press, 1972. However, Morse Library Effectiveness, MIT Press, 1968 and the journal literature were also used extensively.

Eleven students were enrolled in the course. Five were students from the Graduate School of Library Science, four were from the College of Engineering, and two were special students. However, their backgrounds were not distinct with much overlap in skills among those in library science and engineering.

One of the papers focuses on faculty perception of and desires for machine-readable bibliographic information services while the other two papers focus on the problem of providing the document given that you know it exists. This includes consideration of optimizing a closed-stacks system with respect to requestor waiting time and measuring the effect of the library system's geographical dispersion on document retrieval time.

The success of these projects was due, in part, to the cooperation received from the people in charge of the systems being studied. Instead of acknowledging them in this Forward, they are noted in the papers that follow.

Urbana, Illinois
June 1975

William B. Rouse

TABLE OF CONTENTS

ASSESSMENT OF FACULTY INTEREST IN AN INFORMATION RETRIEVAL SERVICE

K.D. MacLaury, S.H. Rouse, L.F. Selander

ABSTRACT

This paper presents the results of a questionnaire survey investigating interest or potential demand for a computer-based information retrieval service. Selection of a stratified random sample of faculty of the University of Illinois at Urbana-Champaign is described. Of the 1040 questionnaires mailed, 434 (42%) were returned and provide data for statistical analysis.

Hypotheses are tested which state the correlation between certain user characteristics and estimated interest in an information retrieval service. An "interest index" is computed from a combination of various responses on the questionnaire. Appropriate parametric and non-parametric statistical tests are used in analyzing the data. Computations and data processing were handled by running the data through SPSS (Statistical Package for the Social Sciences).

In general, this report concludes there is a high degree of interest among the responding faculty of the University. Fifty per cent of the faculty indicated they would use the service frequently (more than 3 times per year) and 46% said they would use the service occassionally (1 to 2 times per year). Significant positive correlations were found between interest in an information retrieval service and use of such information sources as abstracts and indexes and journals, and between interest and the respondent's prior use of an information retrieval service.

INTRODUCTION

The increasing rate of publication of scientific literature makes it difficult for the researcher to quickly and easily identify relevant journal articles and reports. The interdisciplinary nature of some research also makes it difficult to identify relevant journal titles.

A computer-based information retrieval service offers the advantage of rapid and methodical searching of the literature. The user makes decisions about the relevancy of retrieved items without having to spend hours manually searching thousands of references in secondary sources.

The Information Retrieval Research Laboratory (IRRL) of the University of Illinois is in the beginning stages of providing a computer-based information retrieval service for the faculty. Through remote access of on-line information retrieval centers which process the machine-readable versions of many secondary information sources, IRRL can provide the faculty with literature searches covering some aspect of nearly every major discipline.

The main purpose of this survey is to provide IRRL with some measure of faculty interest in an information retrieval service. The results of this survey identify user characteristics which estimate or predict degree of interest and describe the respondents' preference for source of funding.

Selection of a pretest group is described in the following section. Results are reported with respect to the questionnaire design and return rate. Criteria for the selection of a stratified random sample of faculty are also described.

Descriptive statistics (frequencies and percentages of those responding) are presented for each of the variables in the questionnaire. Relationships between variables are examined through crosstabulation and correlation. Computation of an interest index is described and correlated with certain user characteristics. The final section presents a summary of the conclusions based on this survey as well as some recommendations for IRRL.

PRETEST

The main purpose for conducting a pretest in this survey was to check the ambiguity of the questionnaire. We wanted to identify a pretest group which would yield a relatively high response rate and be fairly representative of most departments on campus.

The Mechanical and Industrial Engineering Department was chosen as the pretest group. We felt they would probably be cooperative in responding since they would be informed that the questionnaire was being distributed in partial fulfillment for a Mechanical Engineering Department course. The size of the Department (42) would hopefully represent enough diversity in criticizing the questionnaire.

The questionnaire was revised many times before sending out the pretest version. Within 8 working days of the mailing, all of the pretest returns were received, accounting for a return rate of 42% (19/42). No major changes were made for the final questionnaire but a few multiple choice answers were expanded for more options in responses.

SAMPLE

The sample was selected according to the following criteria:

1. Representative of University of Illinois departments
2. Faculty with official teaching/research responsibilities
3. Faculty ranks within the sample to remain proportional to the population distribution
4. Random selection of the sample according to these criteria.

We also wanted the survey to deal with a relatively large set of data and we wanted about 400-500 returns.

The mechanics of selecting our sample was made relatively easy. Through the Office of Administrative Studies, John E. Terwilliger produced IBM cards corresponding to the population we specified. After selecting our sample from these cards, he produced gummed address labels.

The population of faculty that we were interested in totaled 2125, minus the pretest group which resulted in 2083. We took half the total, yielding a sample size of 1040. Based on the return rate of the pretest, we might expect 437 returns. This reinforced our decision to sample half the population.

The population was arranged by rank within departments and randomly within rank. From this arrangement we pulled every other card to receive a questionnaire. Split appointments were allotted to departments holding the highest percentage, with ties distributed at random to the first department listed for that individual. This procedure elimimated the possibility that some staff members be sent more than one questionnaire. The stratified sample was thus representative of the faculty by ranks within departments.

The final version of the questionnaire was mailed to the sample on April 23. As of May 30 (27 working days) we received 434 responses, or a 42% return rate.

CHOICE OF STATISTICAL TESTS

Since the bulk of our data was either nomical or ordinal, we primarily used non-parametric statistics for the analyses we chose to make. These statistics were the Chi-Square Contingency Test for Goodness of Fit and the Kendall tau b and Kendall tau c methods of computing correlation coefficients. Kendall tau b was used when the number of rows in a contingency table equalled the number of columns; Kendall tau c was used with unequal rows and columns [1].

The parametric statistics used were the Analysis of Variance, the t-test, and the Pearson-product-moment correlation [2].

QUESTIONNAIRE

The main purpose of this questionnaire was to survey the faculty's perception and desire for an information retrieval service with the ultimate goal of providing IRRL with inofrmation about user characteristics which might predict potential interest in the service.

The questionnaire was designed to measure characteristics in three general categories:

1. Individual characteristics
2. Information sources and experience
3. Interest in information retrieval services.

The reasons for including specific questions are discussed in this section as well as some interesting frequencies and correlations. A copy of the questionnaire reporting the percent of returns for each answer is appended to this paper. In order to encourage a high return rate, the questionnaire was limited to one page.

The first general category, individual characteristics, is covered by questions 1 through 3. Identifying the respondents' major academic department provides a convenient basis (department and college) for creating homogeneous subgroups. Department and college information also identifies two levels of management which are potential sources for funding an information retrieval service.

With question 2 we gathered data to test the hypothesis that younger faculty would account for a higher percent of the returns than older faculty. Age and experience variables are measured by rank, highest degree received and year of highest degree. As we expected, year receiving the highest degree correlates negatively with rank. Assuming that the younger faculty received their highest degree most recently, then we can identify younger faculty by rank. Our results support our hypothesis and show that the returned questionnaires are accounted for by 50% of the assistant professors surveyed in the sample, 40% of the associate professors, and 37% of the full professors.

We asked respondents in question 3 to characterize their position by percent time spent teaching, research, and other (usually administration). One hypothesis we wanted to test was that interest in an information retrieval service would correlate positively with research. We felt that researchers were more likely than teachers or administrators to be involved in reporting and collecting current published results.

The second general category, information sources and experience, is covered by questions 4 through 9. A bibliographic information retrieval service such as that offered by IRRL, accesses data bases which primarily cover journal articles, technical reports, and books or monographs. We hypothesized that if a respondent rated these sources very useful he would more likely be interested in an information retrieval service than the respondent who indicated a rating of not useful. Indexes and abstracts were added for the respondents' ratings because they closely correspond with computer-based information retrieval center services. Colleagues and preprints were additional sources mentioned by Garvey in a study of the information gathering habits of research scientists [3]. In general, we hypothesized that a respondent's current use of particular information sources would correlate positively with interest in an information retrieval service.

Correlating percent time teaching and research with rating of information sources (question 4) resulted in a positive correlation between research and journals (Pearson correlation = 0.103, p = 0.034). This result suggests that responding faculty of the University of Illinois spending a large percent of their time in research are more likely to find journal literature more useful than faculty who spend less time in research.

Questions 5 through 8 related to the respondents' past experience with information retrieval services. We expected that those respondents with past experience would correlate positively with interest in an information retrieval service. These results are reported later in the section covering the interest index. We were quite surprised to find that a relatively large percent of the respondents had previous experience with an information retrieval service (30%).

One interesting result is that teaching was negatively correlated with past experience (tau c = -0.090, p = 0.003) while research was positively correlated with past experience (tau c = 0.108, p = 0.0005). Based on these results, we can conclude that those who spend most of their time in research activities are more likely to have experience using an information retrieval service than those who spend most of their time in teaching activities.

The main reason for including question 6, satisfaction with previous experience, is that we expected a positive correlation between satisfaction and interest in an information retrieval service. These results are reported in the section covering the interest index.

Awareness of IRRL was asked in question 9 for two reasons. We expected to find a positive correlation between awareness and interest in an information retrieval service. These results are reported later in the section covering the interest index. IRRL was also interested in finding out the number of respondents who were aware of their service.

INTEREST INDEX

The central parameter of the survey and the most difficult to measure was the amount of interest the respondents had in using an information retrieval service. It was especially difficult since most of the faculty are unfamiliar with information retrieval services. We approached this parameter from several different angles in questions 10 through 13.

We asked respondents what type of service they would like to use, how often they thought they would use it, and how they thought it should be paid for. If respondents indicated they were not interested, they were asked to indicate their reasons.

It was hoped that the brief definitions of current awareness and retrospective searching provided in question 11 would be a sufficient introduction for those respondents unfamiliar with these services.

It should be noted that 27 of the 28 respondents indicating that they would not be interested in using either kind of service also doubted that their area would be covered by data base services. Most of the responses in the "other" category indicated that the respondents felt they were perfectly capable of searching the literature without outside help.

It was felt that source of funding (question 13) was an appropriate measure of interest since checking research grant/contract indicated the respondents' willingness to, in a sense, spend their own money while the other two responses indicated a desire to have somenone else pay for it. This question caused a lot of confusion among the respondents and many wrote comments to the effect that they had no idea as to who should pay for the service.

The responses to these four questions were given various positive or negative weights (depending on whether the response indicated positive or negative interest in an information retrieval service) and the sum of these weights was called the interest index.

The weights were as follows:

Question 10.	Both	4
	Current awareness	3
	Retrospective	2
	Neither	0
Question 11.	Individual responses	-2
	Other	-4
Question 12.	Frequently	4
	Occassionally	1
	Never	-4
Question 13.	Research contract	1
	University	0
	Other	0

In question 10, it was felt that a check for current awareness indicated more interest than retrospective since the current awareness user would use the service more frequently. In question 11, the respondents that checked the "other" category and filled in the blank were felt to be more emphatic in their expression of negative feelings about information retrieval services and thus received a larger negative weight.

As an additional measure of interest, we utilized the fact that the respondents had been asked to give their name and address on the questionnaire if they were interested in receiving a summary of the survey results. It was thought that interest in the questionnaire results would correspond with interest in using an information retrieval service. This proved to be the case since the interest indexes for the two groups (questionnaires with and without return addresses) were significantly different using the t-test (Table 1). Leaving the return address was given a weight of 5. However, in making the t-test comparison, we did not include this 5 as that would have resulted in a meaningless comparison.

Interest Index and Return Address

	Mean	s.d.	*s.e.	N
No Address	5.81	2.70	0.17	259
Address	7.33	1.90	0.16	143

$t = 6.58 \quad p < 0.0005$

*s.e. ≙ standard error ≙ standard deviation/$\sqrt{N}$

Table 1.

Once we had the complete interest index, we could see how it correlated with the variables on the questionnaire. Kendall tau c and Pearson were two different correlation tests available in SPSS that were appropriate for the data we collected (Table 2).

Among the professional characteristic parameters, the most significant correlations with the interest index were with year of highest degree received and percent time spent in research. These, as expected, indicated that researchers and younger faculty are more interested in using an information retrieval service.

In the second group of variables, the high correlations with use of indexes and abstracts and journals were as expected. The strong negative correlation with the "other" category seems to indicate that people who went to the trouble to fill this in have thought about and are probably satisfied with their current information gathering methods. Also, sources covered by information retrieval services were covered by the choices listed on the questionnaire, so people using other less conventional sources would have less use for an information retrieval service.

A strong correlation in this section was with previous use of information retrieval services. One would expect a strong correlation between interest and satisfaction with a previous service (satisfaction and drawbacks variables). However, this was not found. It would seem that once a researcher has tried an information retrieval service, he is anxious to try again no matter how bad the first experience.

Variables Correlated with Interest Index		
Variables	Kendall tau c	Pearson
Rank	-0.092**	----
Highest degree		----
Year received	----	0.162**
Teaching	----	----
Research	0.085**	0.107*
Other	-0.062*	-0.115*
Information sources		
Colleagues	----	----
Journals	0.098**	0.136**
Books	----	----
Technical reports	----	----
Preprints	0.077*	----
Indexes and abstracts	0.264***	0.289***
Other	-0.206*	-0.369*
Experience	0.210***	0.183***
Satisfaction	----	----
Drawbacks	----	----
Awareness	----	----

* ≙ $p < .05$
** ≙ $p < .01$
*** ≙ $p < .001$

Table 2.

RESULTS BY DEPARTMENTS/SCHOOLS/COLLEGES

The mean interest index, standard error and return rate computed by Departments, Schools, or Colleges in rank order is shown in Table 3. (Colleges or Schools were determined by the Accounting Office's code for that group). Since the College of Liberal Arts and Sciences was so large, it was further broken down by the following disciplines: Life Sciences (Botany, Entomology, Microbiology, Physiology and Biophysics, and Zoology); Chemistry (School of Chemical Science, Biochemistry, Chemistry, Chemical Engineering); Social Science (Anthropology, Asian Studies, Geography, Political Science, Sociology); Language Studies (Teaching of English as a Second Language, French, German, Slavic, Spanish, Italian, and Portuguese); Humanities (Classics, Philosophy, History, Linguistics, Comparative Literature, English, Religious Studies); Physical Sciences (Astronomy, Geology); Mathematics; Psychology; Speech; and Liberal Arts and Sciences Administration.

The Schools of Life Sciences and Chemistry are established divisions within the College; the School of Social Sciences is a proposed new school; the other disciplines were determined by us on the basis of our own judgment of what seemed to be logical groupings.

Within the College of Engineering, the Department of Civil Engineering has been reported separately because of its large size and high interest index. The score of 9.42 was the highest in the College (excluding the Mechanical and Industrial Engineering Department which was used for the pretest and therefore not comparable with the other departments). In choosing groups of potential users IRRL might more profitably contact this department singly rather than contacting the College as a whole.

Scores on Interest Index by Departments/Schools/Colleges

Rank Order by Means	Mean	s.e.	N	*Return Rate
1. Labor and Industrial Relations	11.33	1.76	3	67%
2. Basic Medical Sciences	9.50	1.46	8	73
3. Physical Education	9.28	0.95	18	58
4. Environmental Studies	9.00	4.03	2	100
5. Computer Science, Adv. Computation	8.93	0.82	14	57
6. Veterinary Medicine	8.67	1.24	12	41
7. Liberal Arts and Sciences	8.34	0.36	123	37

	Mean	s.e.	N	Rate
Life Sciences	10.71	0.80	14	34
Physical Sciences	9.44	1.27	9	56
Language Studies	9.42	1.10	12	29
Speech	9.10	1.02	10	67
Psychology	9.08	1.20	13	63
Chemistry	8.67	0.92	12	50
Humanities	7.24	0.84	21	25
Social Sciences	7.38	0.78	21	42
Mathematics	4.78	2.05	9	25

	Mean	s.e.	N	*Return Rate
8. Education	8.29	0.60	24	41
9. Engineering (excluding M.E.&I.E.)	8.25	0.54	63	45

	Mean	s.e.	N	Rate
Civil Engineering	9.42	0.89	19	56

	Mean	s.e.	N	*Return Rate
10. Social Work	8.00	0.82	24	41
11. Agriculture	7.89	0.51	63	52
12. Fine and Applied Arts	7.81	0.75	27	25
13. Commerce	7.53	1.13	15	34
14. Aviation	6.67	2.03	3	43
15. Communications	6.40	1.33	5	31
16. Library Science	6.00	----	1	33
17. Law	4.20	1.46	5	46
18. Health, Computer Based Educ.	3.00	----	2	100
GRAND MEAN	8.18	0.20	392	38
Missing cases			42	4
Total returns			434	42

*Based on actual number of returns.

Table 3.

In an attempt to compare the Mechanical and Industrial Engineering Department with the College of Engineering in general, an interest index was computed for Mechanical and Industrial Engineering adjusting weights where necessary and omitting weights for return address, since this was not listed on the pretest. We tested the hypothesis that the pretest displayed the same characteristics as the sample from the College of Engineering. Using the t-test we found the Mechanical and Industrial Engineering Department to have approximately the same mean interest index as the College of Engineering sample (Table 4). The Mechanical and Industrial Engineering Departments's interest index is slightly higher than the College of Engineering and this is due to the slightly higher interest index within the M.E. & I.E. group without past experience. Perhaps the higher interest index for the non-exp[illegible]roup within the M.E. & I.E. Department can be explained by the fact t[illegible]e course motivating this survey was given in their department. Had the M.E. & I.E. Department been included in the sample survey we can conclude that the relative ranking of the College of Engineering in Table 3 would not change.

Comparison of Pretest and College of Engineering	Mean	s.d.	s.e.	N
Mechanical and Industrial Engineering	7.39	1.85	0.44	18
College of Engineering (w/o M.E.&I.E.)	6.19	2.90	0.57	26
$t = 1.67$ $p < .10$				
M.E. & I.E. - experience	6.75	3.86	1.93	4
College (w/o M.E.&I.E.) - experience	6.23	3.24	0.94	12
$t = 0.24$ not significant				
M.E. & I.E. - no experience	7.57	0.94	0.25	14
College (w/o M.E.&I.E.) - no experience	6.17	2.75	0.73	14
$t = 1.80$ $p < .05$				

Table 4.

No a priori hypotheses were made regarding differences in Schools or Colleges' means; however the Analysis of Variance (ANOVA) was performed on the interest index by Schools or Colleges. The College of Liberal Arts and Sciences was considered as a whole. This test was not significant, indicating there are no significant differences between these groups or the tool used for measuring differences is not able to determine any differences.

We also hypothesized that those who had used an information retrieval service would have a greater interest in such a service on campus than those who had not. A t-test comparing the mean interest index of experienced respondents with that of non-experienced respondents is shown in Table 5.

Interest Index Experience/No Experience	Mean	s.d.	s.e.	N
Experienced respondents	9.23	3.82	0.34	125
Non-experienced respondents	7.68	3.92	0.24	265
$t = 3.72$ $p < 0.0005$				

Table 5.

The interest index for experienced respondents in rank order is shown in Table 6. The percent of the respondents with experience is also shown in this table. Again it is interesting to look at Civil Engineering. The mean interest index is 12.13 which is the highest interest index and the largest experienced block within a department.

Rank Ordered Scores of Experienced Respondents on Interest Index

Departments/Schools/Colleges					Mean	s.e.	N	% with exp.
1. Fine and Applied Arts					11.75	1.60	4	15
2. Physical Education					11.12	0.86	8	44
3. Liberal Arts and Sciences					10.25	0.57	28	23
	Mean	s.e.	N	%				
Humanities	13.00	---	1	13				
Psychology	12.40	1.12	5	38				
Life Sciences	11.67	1.02	6	43				
Physical Sciences	11.50	1.32	4	23				
Language Studies	10.00	4.00	2	17				
Speech	8.00	0.82	4	40				
Chemistry	7.33	0.67	3	21				
Social Sciences	6.50	1.50	2	50				
4. Basic Medicine					9.67	1.71	6	75
5. Veterinary Medicine					9.11	1.34	9	75
6. Environmental Studies					9.00	4.00	2	100
7. Engineering (excluding M.E.&I.E.)					8.95	0.89	22	36
	Mean	s.e.	N	%				
Civil Engineering	12.13	0.55	8	47				
8. Education					8.69	0.75	13	54
9. Commerce					8.00	3.06	3	
10. Labor and Industrial Relations					8.00	----	1	25
11. Social Work					8.00	----	1	33
12. Agriculture					7.79	0.95	26	41
13. Computer Science, Advanced Comp.					5.00	----	1	6
Individuals unidentified by dept.							8 *	
GRAND MEAN					9.23	0.34	125	32

* Based on departmental returns.

Table 6.

CONCLUSIONS

In general this report found a high level of interest in the services offered by a computer based information retrieval service. Over 73% of the respondents indicated they would find both current awareness and retrospective services helpful in their professional activities. If not constrained by the cost of the service, 50% of the respondents indicated they would use the service frequently.

Based on the results of this survey, we found three factors which tend to identify potentially interested users of an information retrieval service:

1. Rank
2. Research responsibilities
3. Department.

If the promotional efforts of the Information Retrieval Research Laboratory are focused on any subgroups of the faculty, they should be aimed at younger faculty with predominantly research responsibilities.

With respect to departments we suggest the following criteria as providing useful information for deciding which departments IRRL might contact first:

1. Score on interest index
2. Absolute size of returns
3. Return rate
4. Experience in using an information retrieval service.

The first criterion, score on interest index, is important because it is a measure of the interest a particular department, school or college has in learning about and using an information retrieval service. The second criterion, absolute size of the returns, is important because it indicates which department should yield the greatest number of participants per contact. The third criterion, return rate, could be interpreted as a measure of cooperation indicating perhaps which groups might be easier to work with. This measure is not correlated with the interest index. The fourth criterion is experience. It is important because we found that, in general, groups with experience are more interested in an information retrieval service than those without experience. Also, satisfied experienced users may be a valuable asset in converting skeptical non-experienced members of their departments since they are in a position to alleviate some doubts non-experienced colleagues may have about the value of such a service.

On the basis of these criteria we recommend the following departments and schools as the most likely users of an information retrieval service. Since we are not suggesting any priority within these 9 groups they are listed alphabetically. Naturally, these recommendations are specific to the University of Illinois at Urbana-Champaign and should not be construed as recommendations for universities in general.

Department/School	Interest Index	Size	% Return	Interest Index-Experienced
Basic Medical Sciences	9.50	8	73	9.67
Civil Engineering	9.42	19	56	12.13
Education	8.29	24	41	8.69
Language Studies	9.42	12	29	10.00
Life Sciences	10.71	14	34	11.67
Physical Education	9.28	18	58	11.62
Physical Sciences	9.44	9	15	11.50
Psychology	9.08	13	63	12.40
Veterinary Medicine	8.67	12	41	9.11

ACKNOWLEDGEMENTS

We would like to express our appreciation to the following groups and individuals who contributed to our project. Professor Martha E. Williams, Director of IRRL supported the initial idea of this project and along with the class offered constructive criticism of the questionnaire. John E. Terwilliger of the Office of Administrative Studies offered generous and extensive help in selecting our sample and providing mailing labels. We also appreciate the contribution made by the Mechanical and Industrial Engineering Department for the cost of xeroxing the questionnaire and providing envelopes. Secil Tuncalp of the Survey Research Laboratory offered suggestions for appropriate statistical procedures included in SPSS.

REFERENCES

1. Nie, Norman; Dale H. Bent; C. Hadlai Hull, Statistical Package for the Social Sciences, New York: McGraw-Hill, 1970.

2. McElroy, E.E., Applied Business Statistics, San Francisco: Holden-Day, 1971.

3. Garvey, William D.; Kazuo Tomita; Patricia Wolfe. "The Dynamic Scientific Information User." Information Storage and Retrieval, 10 (No. 3-4):115-31 (1974).

SURVEY RESULTS

1. In which department is your major academic appointment? (If equally divided between departments, please list both departments.)

2. Rank: Asst. 50% 142/286 Prof. 37% 176/482 Assoc. 40% 110/272; Highest Degree: Ph.D. 88.3% Masters 9.7% Other 1.0% Yr. Rec'd: ____

3. How would you characterize your position? (Please circle one number in each category.)

Teaching	0%	25%	50%	75%	100%
	7.3%	23.4%	41.9%	19.2%	8.2%
Research	0	25	50	75	100
	17.8%	32.6%	36.6%	11.5%	1.4%
Other	0	25	50	75	100
	66.4%	20.0%	7.3%	4.0%	2.3%

4. How would you rate the following information sources as to usefulness? (Please circle one number in each category.)

	Very Useful	*Useful*	*Not Useful*		*Very Useful*	*Useful*	*Not Useful*
	41.6%	46.9%	11.5%		30.9%	43.9%	25.2%
Colleagues	3	2	1	*Technical Reports*	3	2	1
	74.6%	24.0%	1.4%		20.3%	45.3%	34.4%
Journals	3	2	1	*Preprints*	3	2	1
	52.0%	43.2%	4.8%		17.1%	51.7%	31.2%
Books	3	2	1	*Indexes & Abstracts*	3	2	1

Other: 3: 87.5% ; 2: 10.0%; 1: 2.5%

5. Have you ever used a computerized information retrieval service(s)?

30.4% *Yes*; 69.6% *No* *(If No, skip to question 9.)*

*6. How satisfied were you with the service(s) you used most frequently?

21.7% *Very Satisfied*; 48.8% *Satisfied*; 29.5% *Not Satisfied*

*7. What were the principal drawbacks to the service(s)? (Please check all applicable answers.)

53.0% *Too much irrelevant material*
25.0% *Too expensive*
42.4% *Important items missed*
9.1% *Liason with service unsatisfactory*
15.9% *Excess delay in getting results*
18.2% *Other:* ____

*8. How was the service paid for? (Please check all applicable answers.)

45.5% *Contracts/Grants*
33.3% *University*
14.4% *Personal*
14.4% *Organization*
11.4% *Other:* ____

*Responses to 6-8 are based on 30% responding YES to question 5.

9. Are you aware that the Information Retrieval Research Laboratory on this campus offers a computer based information retrieval service covering most subject areas?

3.0% *Attended a demonstration/seminar*
71.7% *Not aware*
25.3% *Heard about the service*

10. A computer based information retrieval service provides two modes of literature searching. A "current awareness" service informs users of literature relevant to their subject area covering regular intervals (e.g., monthly). A "retrospective service" covers a longer time span and provides the user with an historical literature search of their field. Which service(s) do you feel would be helpful to you?

74.4% *Both*; 11.5% *Current Awareness*; 7.6% *Retrospective*; 6.5% *Neither*

#11. If neither, why? (Please check all applicable answers.)

1.9% *Do not need to search the literature.*
2.5% *Do not want to be glutted with information.*
0.5% *Current literature in my area is not worth reading.*
6.3% *Doubt my area would be covered.*
0.7% *Would have been interested in such a service earlier in my academic career.*
4.2% *Other:* ____

#Responses to this question based on 6.5% responding NEITHER to question 10.

12. If such a service were available at no cost to you personally, would you use it?

50.4% *Frequently*; 46.6% *Occasionally (1 or 2 times a year)*; 2.8% *Never*

13. Who do you think should pay for such a service?

38.2% *Your research grant/contract*; 77.9% *University*; 7.6% *Other:* ____

MODELING CLOSED STACKS DOCUMENT RETRIEVAL

J. Beal, Y-Y. Chu, J. Greenstein, S. Von Vogt

ABSTRACT

The stacks of the University of Illinois Main Library are closed to most undergraduate students of the University. These students must submit requests for documents to the circulation desk personnel for servicing rather than enter the stacks to locate documents themselves. During periods of heavy undergraduate use, waits of one hour for the servicing of such requests are not uncommon. This paper describes the present operation of the document retrieval system and develops a model for a portion of the system. The collection and analysis of data needed for predicting performance of this subsystem and the design and use of a queueing model that depicts the operation of the subsystem are discussed. The model is then used to predict subsystem performance for various staffing policies and levels of user demand. These predictions can be used to identify the most effective staffing policies available to the manager under the constraints imposed by limited availability of resources.

INTRODUCTION

The stacks of the University of Illinois Main Library are closed to most undergraduate students of the University. These students must submit requests for documents to the circulation desk personnel for servicing rather than enter the stacks to locate documents themselves. During periods of heavy undergraduate use, waits of one hour for the servicing of such requests are not uncommon. This paper describes the present operation of the document retrieval system and develops a model for a portion of the system. The model is then used to predict performance of this subsystem for various staffing policies and levels of user demand.

Documents are shelved on the ten floors (or "decks") of the stacks according to call number. The circulation desk is located on the fifth deck at the only entrance to the stacks. "Pages" are stationed on several of the decks. In addition to having reshelving duties, pages locate documents requested by the circulation desk and dispatch these documents to the circulation desk. Decks on which pages are stationed are termed "open" decks. Users without stacks privileges write and submit request cards for documents to the circulation desk personnel. The deck location of the requested document is determined and the request card is sent by pneumatic tube to the page responsible for servicing requests on that deck.

A page receives document request cards on his open deck. When a request arrives through the pneumatic tube, the page discontinues reshelving and services the request. He attempts to locate the requested document. If the document is found, it is sent with the request card by conveyor to the circulation desk. If the document is not found, the request card is sent to the circulation desk through the pneumatic tube system. Upon completing service on a batch of requests, the page returns to the open deck to continue reshelving or to begin service on any new requests that have arrived in his absence.

As requested items or request cards for items not found arrive at the circulation desk by conveyor or pneumatic tube, circulation desk personnel collect the documents and request cards and complete the service to the waiting patrons. Located documents are charged out and users whose requests could not be found are notified.

When the number of requests to a page becomes large, the amount of time that elapses from the moment the request card arrives at the open deck to the moment the page finishes servicing that request increases. The elapsed time can, during periods of peak demand, reach durations of over one hour. Because such a significant amount of the total service time for document requests is due to the service time of the page, it was decided that the document request-page portion of the document retrieval system should be studied in detail. More specifically, the development of a quantitative model of this portion of the system was deemed desirable. With such a model, the performance of the document request-page subsystem could be predicted for various staffing policies and levels of user demand. These predictions could then be used to formulate and identify the most effective staffing policies available to the manager under the constraints imposed by limited availability of resources.

APPROACH

The document retrieval system can be modeled as a series of queueing systems. Patrons without stacks privileges form queues in front of the circulation desk waiting for service on their requests to be completed by the circulation desk personnel. The patrons are treated as customers and the circulation desk personnel as servers of a queueing system. The circulation desk personnel send document request cards to the pages for servicing. The request cards can now be treated as customers queueing up to be served, while the pages can be treated as the servers of the queue. When the page completes service of a request he returns the request card and document to the circulation desk. The cards and documents arriving at the circulation desk can then be treated as customers in a third queueing system, waiting to be served by the circulation desk personnel. An illustration of how this type of model might describe the operation of the document retrieval system is given in Figure 1.

The queueing model representing the document request-page portion of the system (Figure 2) is the model to be developed in detail, as it is this portion of the system that accounts for much of the total time required to service a request. The requests arriving to a page are customers forming a queue. The page is a server for these customers. The request cards can be characterized by the probability distribution of the time between request arrivals. The page can be characterized by the probability distribution of the time it takes him to service requests. These probability distributions are determined from measurements of the arrival rates and service times that

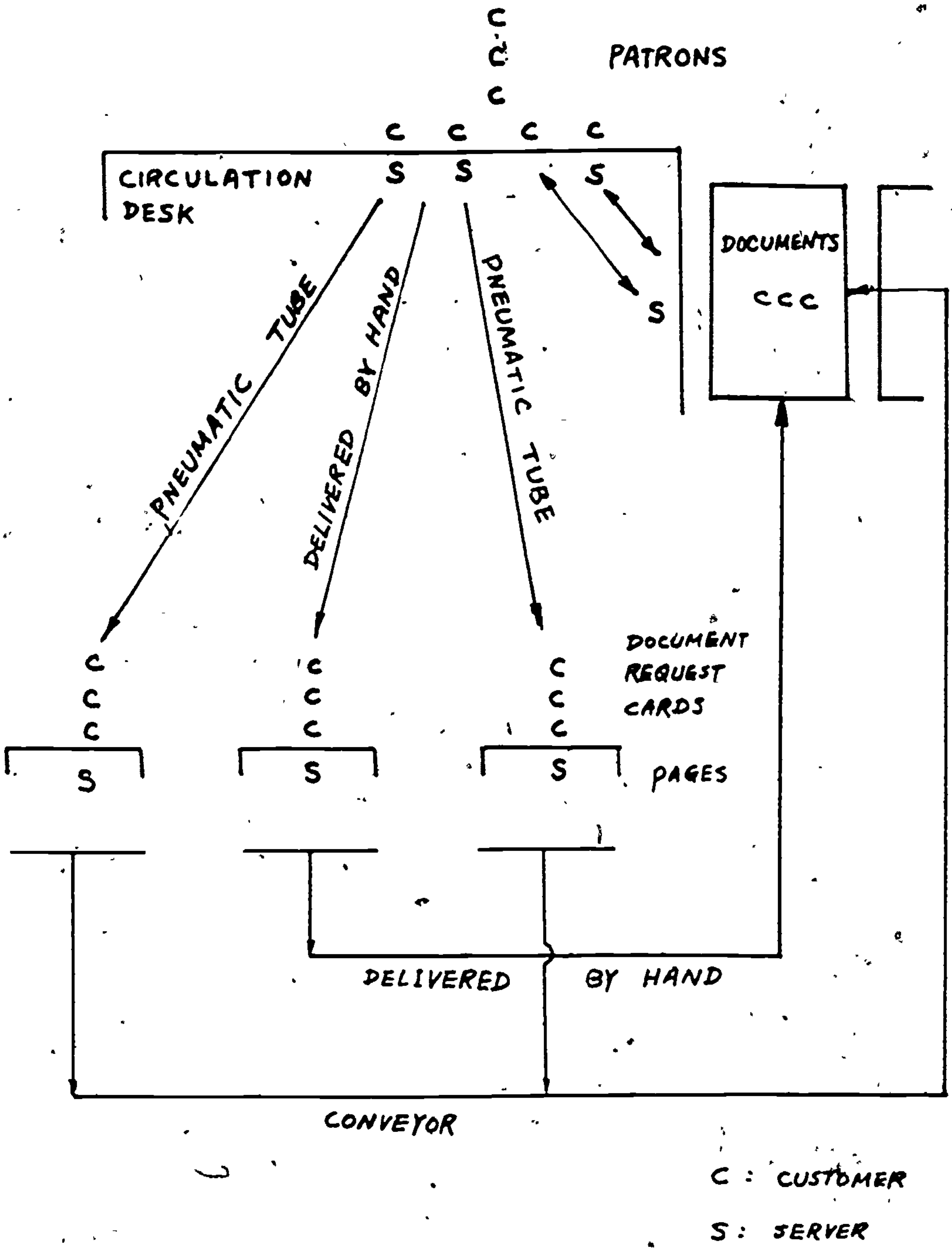


Figure 1
The Document Retrieval System
as a Series of Queueing Systems

actually occur in the stacks system. After proper analysis of this data, the resulting distributions can be used in an appropriate model of the queueing system to yield predictions of the average amount of time a request card must wait for service and the average length of the queue of requests waiting to be serviced. Various staffing policies can then be used in the model and the effects of policy changes on waiting times and queue lengths can be determined.

The two basic steps involved in obtaining the desired information on waiting times and queue lengths are then 1) the collection and analysis of arrival rate and service time data from the document request-page portion of the stacks system; 2) The design and use of a queueing model that accurately depicts the operation of the document request-page portion of the stacks system. The following sections of this paper discuss in some detail the execution of these steps for the document request-page subsystem.

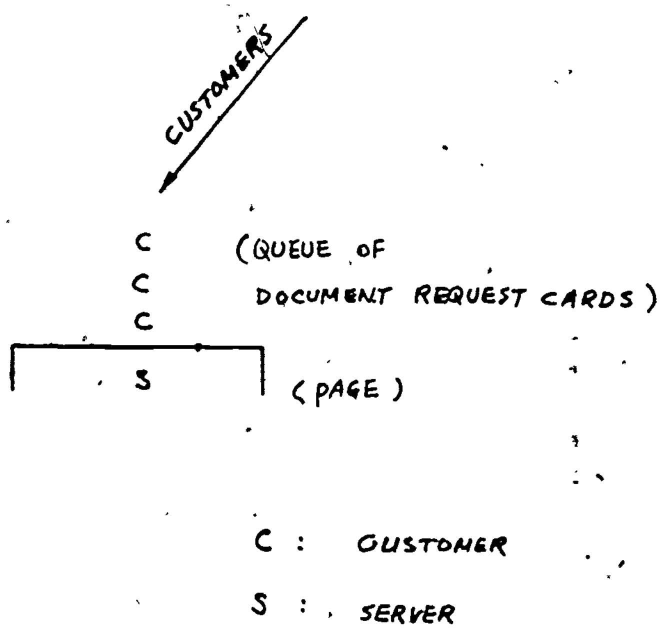


Figure 2
The Document Request-Page Queueing System

COLLECTION AND ANALYSIS OF ARRIVAL RATE DATA

Data on the arrival rates of document requests to each deck of the stacks was collected continuously for nine days. Because of the number of requests, decks, and hours involved, the pages were asked to note request arrivals themselves as they occurred. A data sheet was designed that required only a check mark by the page as each request arrived. The data collected on the sheets could then be used to determine the arrival rate of requests to each deck for each hour of the day and each day of the week. Pages were also asked to enter a check when requested items could not be located. This allowed the determination of the percentage of requests that could not be located for different arrival rates and decks. A copy of a typical data collection sheet is included in the appendix.

The probability distribution of the time between request arrivals could not be determined without requiring the page to accurately note the time of arrival of each request. It was felt that asking this of the pages would degrade their job performance while yielding results of questionable reliability. Morse [1] has shown that the Poisson distribution often accurately represents the statistics of library arrival processes. This implies that the times between request arrivals have an exponential distribution. The exponential distribution has the property that the time until the next arrival is uninfluenced by the time at which the last arrival occurred. This seems a reasonable description of the arrival process of document requests. The interarrival times of document requests were therefore assumed to be exponentially distributed.

The arrival rate data was collected during the twelfth week of the semester, the beginning of a period in which documents located in the stacks are heavily demanded by undergraduates. Figure 3 shows the average number of document requests submitted to the pages for each day of the week. The number of requests tends to be high during the first four days of the week (Monday through Thursday) and tapers off on the weekend. Figure 4 shows the average number of requests submitted to the pages for each hour of the day. The number of requests increases steadily through the morning and early afternoon, peaks from 3 to 4 p.m., declines drastically during the dinner hours and increases again in the evening. Figure 5 shows the average number of document requests to each deck in a day. The decks can then be listed in decreasing order of undergraduate use as follows: 1,3,2,10,6,4,9,5,8,7.

The total number of requested documents for the nine day period of data collection was 3,736 of which 1,383 or 37% were not found. The number of requests not found from 9 to 10 a.m., a slow hour, and from 3 to 9 p.m., the hour of peak demand, were compared to determine whether the percentage of documents not found increased with the number of requests arriving to the pages. The arrival rate from 3 to 4 p.m. is almost four times the arrival rate from 9 to 10 a.m. 31% of the requests arriving from 9 to 10 a.m. were not found while 41% of those

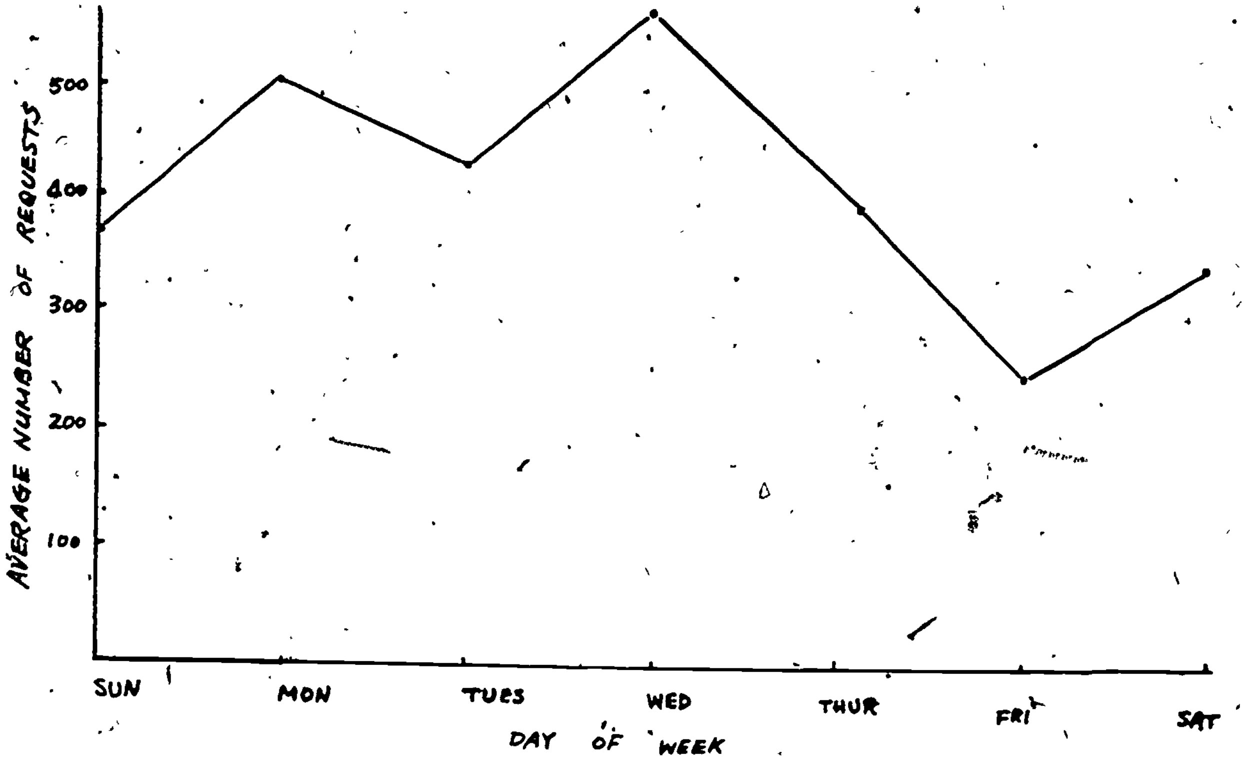


Figure 3
Average Number of Request Arrivals by Day

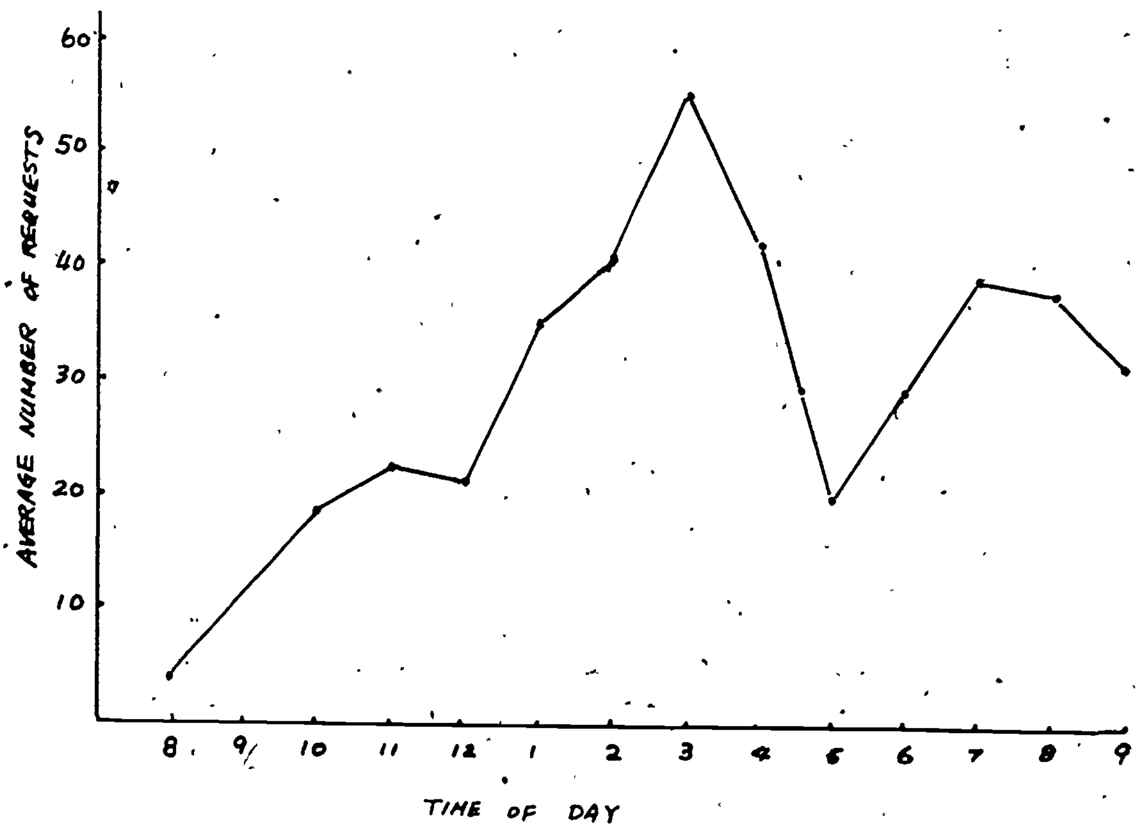


Figure 4
Average Number of Request Arrivals by Hour

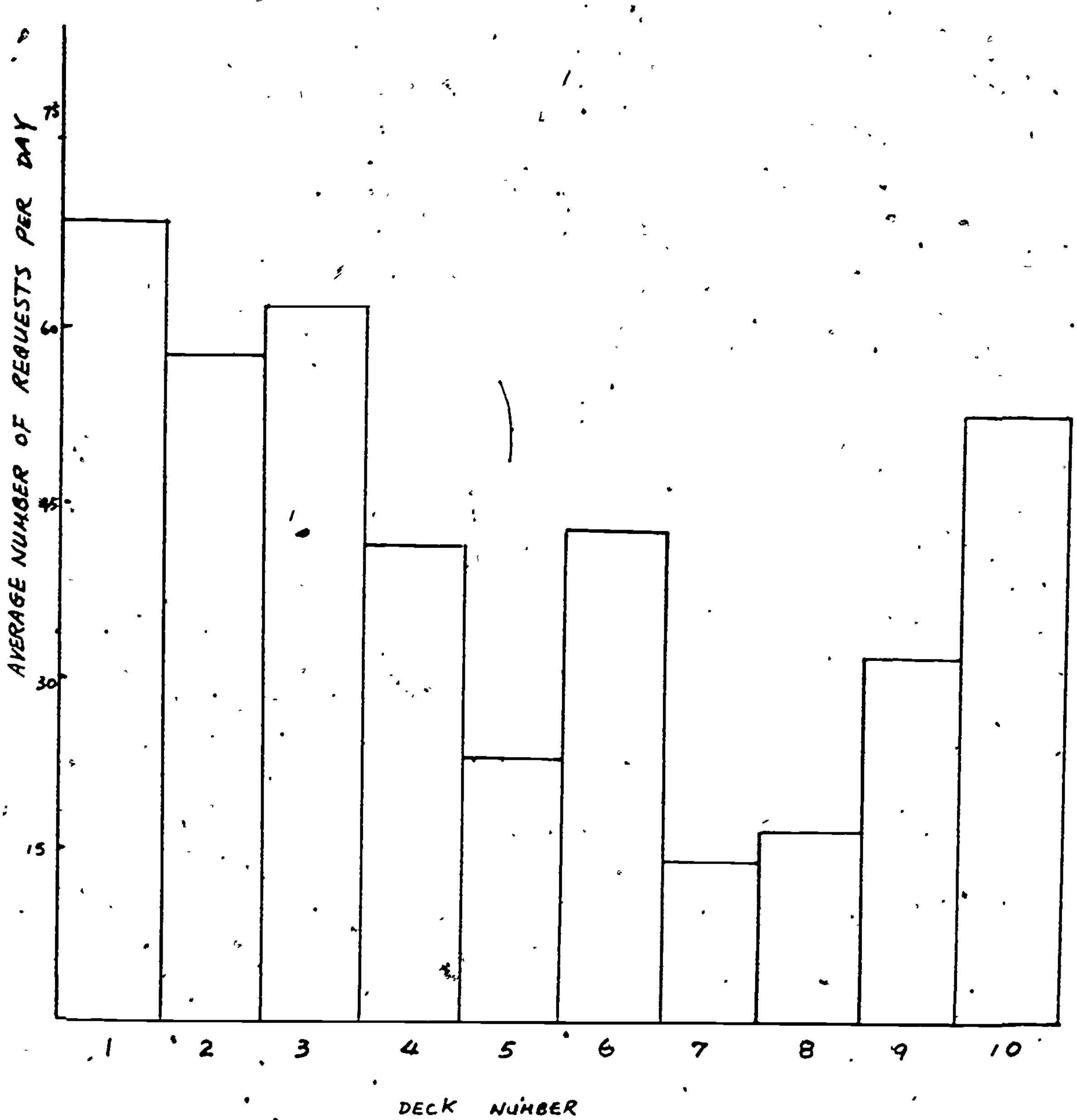


Figure 5

Average Number of Requests Arriving at Each Deck Per Day arriving from 3 to 4 p.m. were not found. The increase in the percentage not found is not statistically significant. This gives some indication that the pages are consistent in their performance for wide ranges of demand.

COLLECTION AND ANALYSIS OF SERVICE TIME DATA

The average amount of time taken by a page to service a document request can be expected to be a function of several variables. In particular the service time is likely to be strongly influenced by the number of requests the page collects before leaving his station to service the requests (the batch size) and by the distance the page must travel to reach the requested item. In collecting the service time data, it therefore becomes important to measure service times for various values of batch size and distance, where distance is represented by some postulated distance measure, such as the number of decks between the open deck and the deck on which the document is located. Where the arrival rate data collection only requires keeping track of the occurrences of requests, the service time data collection requires measurement of the time elapsing from the initiation to the completion of service on a request as a function of batch size and distance. Because of the complexity of the measurement task, stacks personnel clearly could not be expected to measure the service times accurately while also performing their assigned duties. Service time data, then, had to be measured by people dedicated solely to this task and was therefore expensive to collect. Additionally, the data could not be collected by watching the pages as they serviced requests, because the constant presence of observers armed with stopwatches would quite probably affect their performance.

The data was collected by several workers (the authors of this paper) all located at the circulation desk. The time at which document requests were sent to an open deck was noted along with the number of requests in the batch, the location of the open deck and the location of each requested document. The time at which each document arrived by conveyor at the circulation desk (or, for documents which were not found, the time at which the request card returned by pneumatic tube to the circulation desk) was also noted. Measurements of service time from the circulation desk had the advantage that 1) the pages were unaware of the presence of data collectors and their service was therefore not affected by this presence, and 2) the requests being submitted by users to the circulation desk could be grouped into batches of various sizes and having various distance measures before being dispatched to the page. Requests could also be created artificially by the data collectors to measure service times for batch sizes or distance measures that would not have otherwise occurred during the data collection. A disadvantage of data collection from the circulation desk was that the presence of several data collectors intercepting request cards interfered with the work of the circulation desk personnel. In fact, data could only be collected during the "slow" hours of the day when the resulting interference was minimal. It should be noted that data collection from the circulation desk also introduces an additional complication - a batch of requests to a page must be held at the circulation desk until the page has completed service on the previous batch. Otherwise, the service times measured for the second batch would include an unknown portion of the service times measured for the first batch.

Service time data was collected on two successive Saturday nights from 7 p.m. until 10 p.m. Service time data collection by its very nature involves a great deal of time, and with the complexity introduced by the necessity of noting times at which request cards are sent and returned, documents arrive by conveyor and users arrive with more request cards (events which often seem to occur simultaneously), an investment of 24 man hours resulted in the collection of service times for only 32 batches of requests.

The service time data was collected for various values of batch size and distance, where distance was represented by a measure of the number of decks from the open deck to the deck on which the requested item was located. The amount of time needed by a page to complete service on a batch of requests can be expected to increase with increasing batch size and distance travelled. A reasonable expression for the expected value of the service time on a batch of requests might then be:

$$E(t_s/B,D) = t_o + C_1B + C_2D$$

Where $E(t_s/B,D)$ is the expected value of the service time, t_s, on a batch of requests having batch size B and distance measure D and t_o, C_1, and C_2 are constants determined from the data by a linear regression technique such that the resulting linear function best fits the service time data. Using the service time data collected, these constants took values of

t_o = 3.3 minutes
C_1 = 1.75 minutes/request
and C_2 = 0.67 minutes/deck

The expected value of the service time on a batch of requests as a function of batch size B and distance measure D can then by expressed as:

$$E(t_s/B,D) \doteq 3.3 + 1.75\ B + 0.67\ D$$

The function fits the data with a standard deviation of 1.9 and is reasonable within the limits of the service time data collected. A much larger collection of service time data is necessary, however, to determine with confidence the relationship between service time, batch size, and distance. More data is also needed to determine with confidence the probability distribution of the service times. The exponential distribution again seems to describe the situation reasonably well. It was therefore assumed that the service times were exponentially distributed.

THE DESIGN AND USE OF A QUEUEING MODEL SIMULATION PROGRAM

With the appropriate data on arrival rates and service times collected and available, the remaining work involves formulating and using a queueing model that accurately depicts the operation of the document request-page portion of the stacks system. The manner in which the page services requests, collecting them in batches and completing service on each request one by one as he drops located documents on the conveyor (or returns cards for unlocated items) has been described. There are, unfortunately, no analytical solutions available for this type of queueing system. A simulation of the system on a digital computer becomes the most attractive approach to analyzing the system.

The total amount of time a request card waits in queue including the time actually spent servicing the request is the variable of interest. The arrival rate and service time data used in a reasonably descriptive simulation of the queueing system should yield as output the average waiting time for a request (including service time) for various staffing policies. The simulation program must, given the interarrival time and service time distributions, predict the time at which an event will occur and determine whether this event is the arrival of a request to the page or the completion of service on a request by the page. As the program generates events and determines the times at which they occur, it keeps statistics on the average waiting time for the requests. After simulating thousands (or millions) of events on the basis of the interarrival and service time distributions, the value of the average waiting time can be expected to converge to the actual value to the extent that the simulation program describes the operation of the actual system.

Several assumptions were made in writing the simulation program and the simulation describes the actual system to the extent that these assumptions about the operation of the system are true:

1. The arriving requests are assumed to enter the queue and to wait for servicing by a page without reneging (leaving the queue before being serviced).
2. The interarrival times and service times are assumed to be exponentially distributed.
3. The page is assumed to follow a specific operational routine. When locating requested items he travels to the highest deck on which a requested item is located and works successively downward from this deck. He is assumed to dispatch documents by conveyor from the deck on which the items are located. He is finally assumed to return to the open deck upon completion of service on a batch of requests.
4. The proportion of the total number of requests arriving to each deck is assumed to be accurately reflected by the arrival rate data.

5. Each page is assumed to be responsible for a section of the stacks, but no two pages service the same section. That is, there is no overlapping of page responsibilities. The simulation program is run independently for each page and set of deck responsibilities.

The simulation program then requires the following input.

1. The mean value of the request interarrival time for the case being studied. (The interarrival times are assumed to be exponentially distributed about this mean.)
2. The mean value of the service time on a request as a function of batch size and distance travelled. (The service times are assumed to be exponentially distributed about this mean.)
3. The proportion of the total number of requests arriving to each deck. (This is determined from the arrival rate data.)
4. The deck on which the page is stationed (the open deck).

The relevant output of the program is the average waiting time for a request (including service time) under the conditions represented by the input data. A flow chart of the simulation model and a program listing are included in the appendix.

RESULTS

The simulation program was used with the appropriate arrival rate and service time data to obtain the results illustrated in Figures 6 and 7 and Table 1. Each simulation run represents an aggregation of the events occurring over 10,000 simulated minutes - the events being request arrivals and service completions. Figure 6 plots the arrival rate data by deck at 9 a.m. It also gives the waiting time that results from selecting each of the decks as the open deck for a single page working during this hour. It can be seen from the figure that the optimal open deck is generally close to a deck on which many requested items are located. The proportion of requests for items on each deck and the dispersion of the requests over the decks are the factors which strongly influence the optimal deck location. The assumption that the page follows a specific operational routine in which he travels to the highest deck on which a requested item is located and works successively downward tends to favor the selection of higher decks as optimal locations for the open deck.

Figure 7 gives the average waiting time at 9 a.m. and 3 p.m. for various numbers of pages. For a given hour of the day and a given number of pages, all possible assignments of deck responsibilities to the pages can be simulated (subject to the constraint that the pages' deck responsibilities do not overlap). The waiting times given in Figure 7 correspond to the choice of deck responsibilities that results in the smallest waiting times. The 9 to 10 a.m. hour is the slowest hour of the day while the 3 to 4 p.m. hour is the busiest. Curves for other hours of the day can be expected to fall in between the two curves given. Figure 7 shows that from 9 to 10 a.m. the

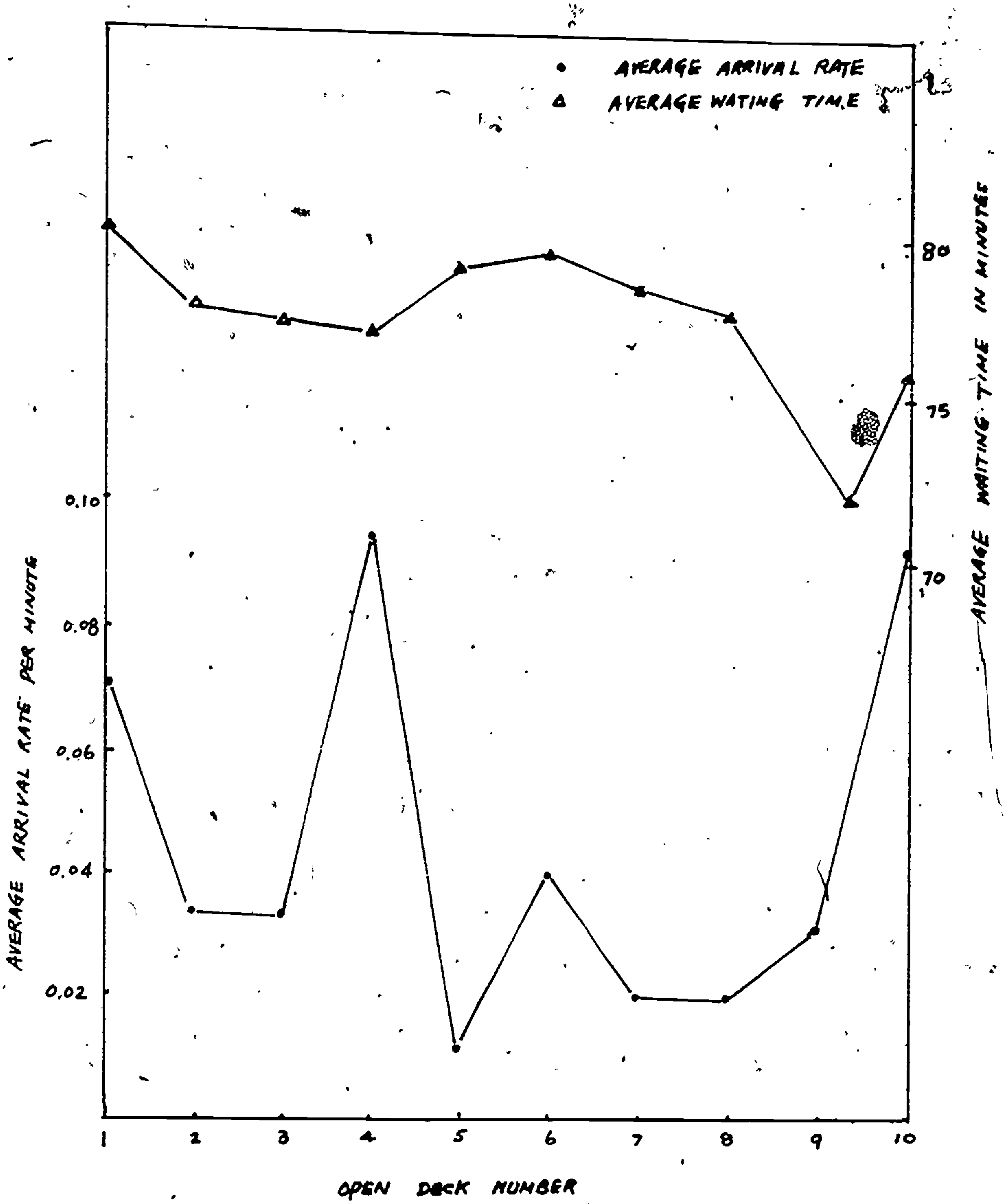


Figure 6
Average Arrival Rate and Average Waiting Time Versus Number of Open Deck for One Page at 9 a.m.

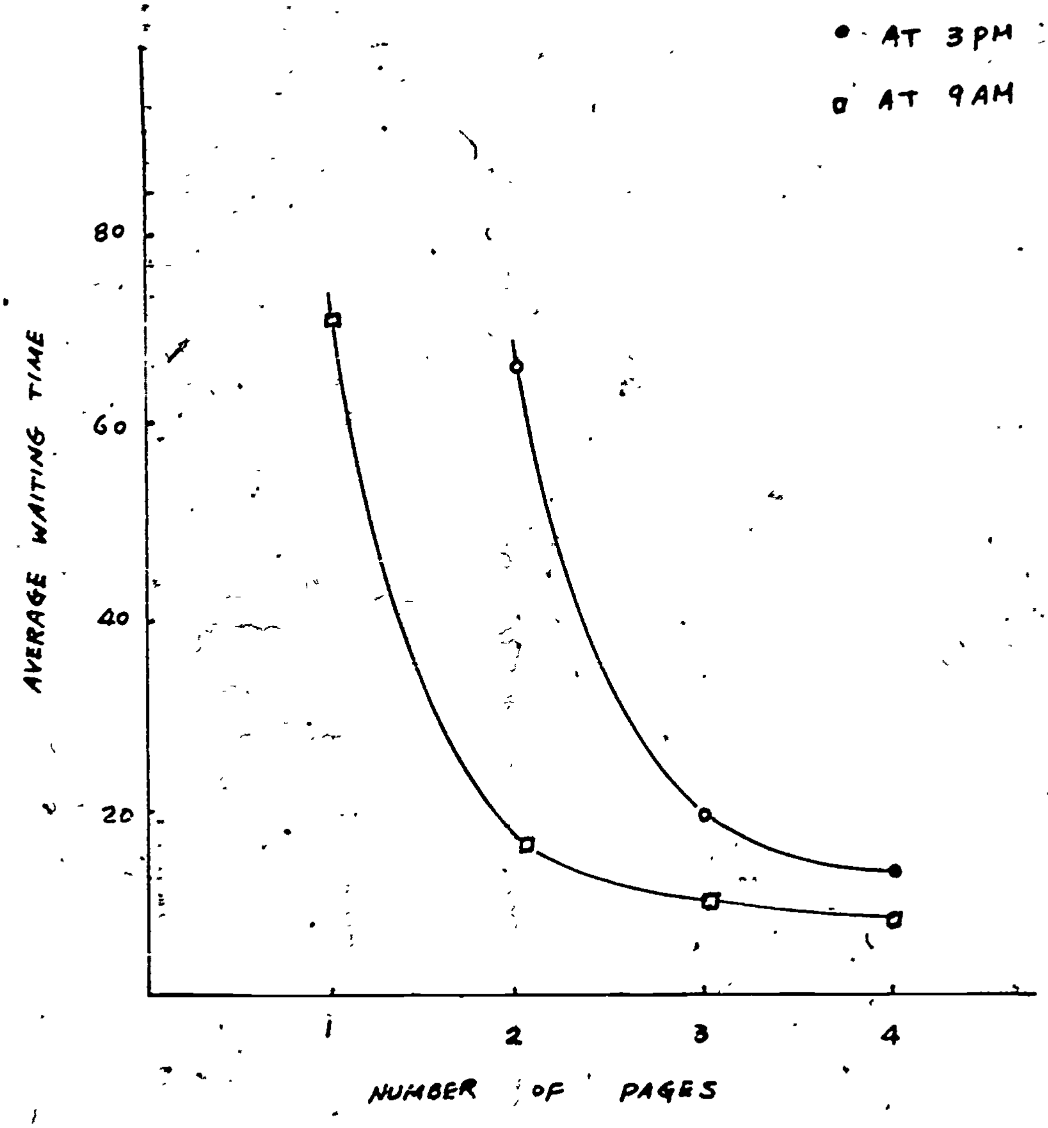


Figure 7
Average Waiting Time at 9 a.m. and 3 p.m.
Versus Number of Pages in the System

TIME PERIOD IN A DAY	NUMBER OF PAGES	OPTIMAL OPEN DECK AND RESPONSIBILITIES	AVERAGE WAITING TIME (MIN.)
9AM-10AM	1	9(1-10)	71.9
9AM-10AM	2	4(1-5) 10(6-10)	15.5
9AM-10AM	3	4(1-4) 10(6-10) 5	14.8*
9AM-10AM	3	2(1-3) 4(4-7) 10(8-10)	10.6
9AM-10AM	4	1(1-2) 4(3-5) 6(6-8) 10(9-10)	8.7
3PM-4PM	2	4(1-5) 9(6-10)	66.6
3PM-4PM	3	2(1-4) 9(6-10) 5	59.2*
3PM-4PM	3	2(2-1) 6(3-6) 10(7-10)	19.3
3PM-4PM	4	1(1) 3(2-3) 6(4-7) 10(8-10)	13.1

*EVALUATED UNDER THE CONDITION THAT DECK 5 MUST BE ONE OF THE OPEN DECKS WITH THE PAGE ON THAT DECK SERVICING ONLY THAT DECK.

Table 1
Simulation Results for Optimal Allocation and Performance

document retrieval system can be operated almost as efficiently with two pages as with three, while operation with one page would result in significantly degraded performance. From 3 to 4 p.m., increasing the number of pages from two to three results in a large decrease in waiting times (from almost 70 minutes per request to 20 minutes per request). No point is plotted for one page from 3 to 4 p.m. as the simulation indicates that in this situation the document retrieval system is unstable. The request arrival rate is greater than the pages' service rate and the number of requests queueing up in the system increases without bound under such conditions.

Table 1, in addition to giving the optimal waiting time that results for a given number of servers for the hours of 9 to 10 a.m. and 3 to 4 p.m., also lists the deck responsibilities and open decks that result in these optimal waiting times. The effect of the additional constraint that one page be stationed on deck 5 servicing only that deck is also illustrated. From 9 to 10 a.m., with 3 pages working, this results in an increase in waiting time to 14.8 minutes per request from the 10.6 minutes per request of the optimal solution. From 3 to 4 p.m., with 3 pages working, the waiting time increases to 59.2 minutes per request from the 19.3 minutes per request of the optimal solution.

Figures 6 and 7 and Table 1 represent only a portion of the results that can be obtained using the queueing model simulation. Using appropriate arrival rate data, the effects of changes in staffing policy for different days of the week or periods of the year can also be studied. The effects of policy changes can be compared in terms of the resulting values of waiting time experienced by arriving document requests.

CONCLUSION

The queueing model simulation of the document request-page subsystem yields a measure of performance, the average time taken by the page to complete service of a request, that can be used to identify the most effective staffing policies available to the manager in terms of that measure of performance. With the proper arrival rate data and sufficient service time data, the model can be used to determine the changes in waiting times that result from adding pages to or subtracting pages from the system at given hours of the day, days of the week, or weeks of the year. The open deck and deck responsibilities for each page that yield the most effective performance in terms of waiting time can also be determined.

Service time data was found to be expensive to acquire. Because of time limitations, the service time distribution used in the system simulation was assumed on the basis of a small amount of data. The reliability of the predictions given by the simulation program could be increased by the acquisition and analysis of greater amounts of service time data. A similar statement holds for the arrival rate data, although it was found to be much less expensive to acquire. Data on arrival rates was taken for one week. If staffing policies for different periods of the year are to be studied, arrival rate data for each of the different periods should be taken. If the effect of additional pages during peak hours of the day is to be examined, the reliability of the resulting predictions can be increased by acquiring more data on arrival rates during the peak hours.

It should again be noted that the simulation program, as it is presently written, does not allow for overlapping page responsibilities. That is, the program assumes that each page is responsible for a section of the stacks, but no two pages service the same section. It may well be that waiting times can be reduced still further by allowing more than one page to work from an open deck or to be responsible for the same section of the stacks. It might therefore by useful to further develop the simulation program so that the effects of such staffing policies can be studied.

The model developed in this paper describes the performance of the document request-page portion of the closed stacks document retrieval system. This portion of the system was modeled in detail because it accounts for a large part of the total time required to service a request. In the future, the subsystems involving the patron-circulation desk personnel interface and the document-circulation desk personnel interface might be studied. The three subsystem models might then be linked in a computer simulation to determine overall system performance as a function of various variables of interest. The effects of adding additional employees at the circulation desk or of changes in the responsibilities of the individual circulation desk employees, for example, could be studied. The performance of the entire closed stacks document retrieval system, rather than the performance of the document request-page subsystem alone, could then be optimized.

REFERENCES

1. Morse, P.M., Library Effectiveness, Cambridge, MA: MIT Press (1968).

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support and enthusiasm of Jean Seyfarth, Circulation Librarian, Stella Mosborg, Assistant Circulation Librarian, Karen Streuss, Bookstacks Librarian, and William Hannaford, Jr., Circulation Assistant, all at the University of Illinois at Urbana-Champaign.

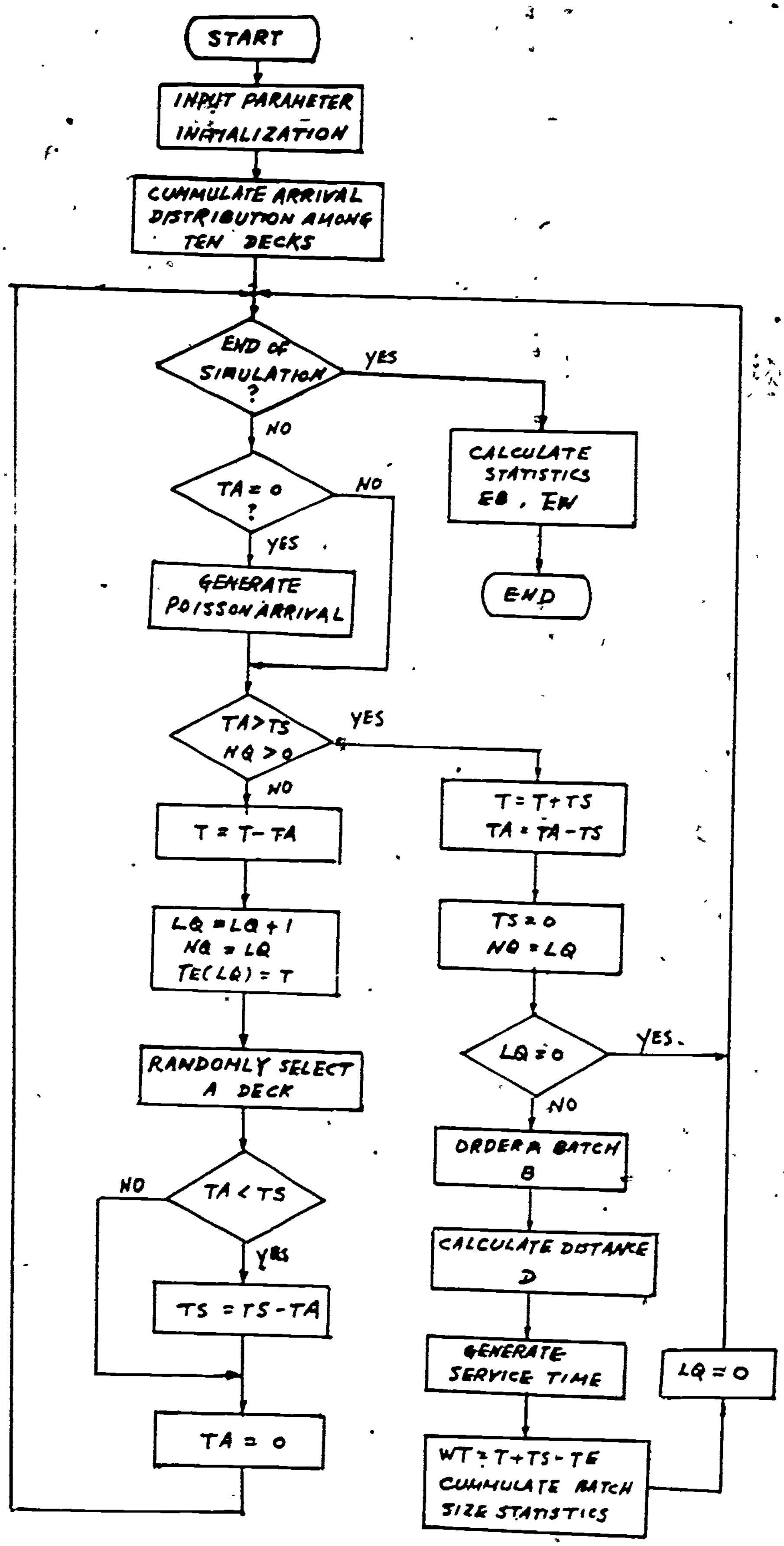
The simulation program presented here is an adaptation of one developed in conjunction with a research project entitled A Mathematical Model of the Illinois Interlibrary Loan Network which is funded by the Illinois State Library and being carried out by the Coordinated Science Laboratory at the University of Illinois.

APPENDIX FOLLOWS

DATE: ____________ OPEN DECK: 5

INITIALS														
TIME	8	9	10	11	12	1	2	3	4	5	6	7	8	9
DECK TEN														
DECK NINE														
DECK EIGHT														
DECK SEVEN														
DECK SIX														
DECK FIVE														
DECK FOUR														
DECK THREE														
DECK TWO														
DECK ONE														

Arrival Rate Data Sheet



EB. ESTIMATED BATCH SIZE

EW: ESTIMATED WAITING TIME

LQ · NUMBER IN QUEUE

T: TIME COUNTER

TA - ARRIVAL TIME

TE · TIME OF ENTERING QUEUE

TS : SERVICE TIME

Simulation Flow Chart

```
C          QUDECK
C
C          SIMULATES CLOSED STACKS DOCUMENT RETRIEVAL
C
C          ME 393/LS 450  MAY 4, 1975
C
           DIMENSION TARR(200),IDECK(200),TIME(200),ARTE(10),P(10)
C
C          INPUT SIMULATION PARAMETERS
C
100        DO 400 I=1,10
           WRITE(5,350)I
350        FORMAT(1X,'REQUESTS/MINUTE FOR DECK ',I2,' = ',$)
400        READ(5,550)ARTE(I)
           WRITE(3,420)(ARTE(I),I=1,10)
420        FORMAT(5F)
450        WRITE(5,460)
460        FORMAT(1X,'DECK0 = ',$)
           READ(5,470)IDECK0
470        FORMAT(I)
           IF (IDECK0)1800,100,480
480        CONTINUE
550        FORMAT(F)
           TLOW=100
           TEND=10100
C
C          INITIALIZE TIME, QUEUE LENGTH, AND SUMS OF DATA
C
           T=0.0
           TA=0.0
           TS=0.0
           LQ=0
           WSUM=0.0
           WSUMS=0.0
           FNW=0.0
           BSUM=0.0
           BSUMS=0.0
           FNB=0.0
           ARATE=0.0
           DO 700 I=1,10
700        ARATE=ARATE+ARTE(I)
           PREF=0.0
           DO 750 I=1,10
           P(I)=PREF+ARTE(I)/ARATE
750        PREF=P(I)
C
C          START SIMULATION LOOP
C
800        TS=0.0
           NC=LQ
C
C          GENERATE SERVICE TIME
C
           IF (LQ.EQ.0) GO TO 1200
C          [illegible] BATCH
           DO 875 I=1,LQ
           [illegible]
           DO [illegible] J=1,LQ-1
           IF (I[illegible]CK(J).[illegible].IDECK(J+1)) GO TO 850
           I[illegible]=1
           [illegible](I)
```

```
      ITEMP=IDECK(J)
      TARR(J)=TARR(J+1)
      IDECK(J)=IDECK(J+1)
      TARR(J+1)=TEMP
      IDECK(J+1)=ITEMP
850   CONTINUE
      IF (ISW.EQ.0) GO TO 890
875   CONTINUE
890   CONTINUE
      IREF=IDECK0
      TREF=3.3
      DO 900 I=1,LQ
      DIST=IDECK(I)-IREF
      SRATE=1.75+0.67*ABS(DIST)
      SRATE=1.0/SRATE
      TIME(I)=TREF+ALOG(RAN(DUMMY))/(-SRATE)
      IREF=IDECK(I)
900   TREF=TIME(I)
      RETDIS=IDECK(LQ)-IDECK0
      TS=TIME(LQ)+0.67*ABS(RETDIS)
C
C     CUMULATE BATCH SIZE STATISTICS
C
      IF (T.LT.TLOW) GO TO 1200
      FLQ=LQ
      BSUM=BSUM+FLQ
      BSUMS=BSUMS+FLQ**2
      FNB=FNB+1.0
C
C     FINISH SERVICE OF LAST BATCH
C
      DO 1100 I=1,LQ
C
C     CALCULATE TOTAL WAITING TIME
C
      WT=T+TIME(I)-TARR(I)
C
C     CUMULATE WAITING TIME STATISTICS
C
      WSUM=WSUM+WT
      WSUMS=WSUMS+WT**2
1100  FNW=FNW+1.0
1200  LQ=0
C
C     CHECK FOR SIMULATION END
C
      IF(T.GE.TEND) GO TO 1600
C
C     GENERATE A POISSON ARRIVAL
C
1300  IF(TA.GT.10.0E-6) GO TO 1400
      TA=ALOG(RAN(DUMMY))/(-ARATL)
1400  IF(TA.GE.TS.AND.NQ.GT.0) GO TO 1500
      T=T+TA
      LQ=LQ+1
      NQ=LQ
      TARR(LQ)=T
C     RANDOMLY SELECT A DECK
      R=RAN(DUMMY)
      DO 1453 I=1,10
      IF (R.GT.P(I)) GO TO 1450
```

```
      IDECK(I5)=I
      GO TO 1475
1450  CONTINUE
1475  CONTINUE
      IF (TA.LT.TS) TS=TS-TA
      TA=0.0
      GO TO 1300
1500  T=T+TS
      TA=TA-TS
      GO TO 840
C
C     CALCULATE STATISTICS
C
1600  EB=BSUM/FNB
      EW=WSUM/FNW
      SB=SQRT((BSUMS-FNB*EB**2)/(FNB-1.0))
      SW=SQRT((WSUMS-FNW*EW**2)/(FNW-1.0))
C
C     OUTPUT
C
      WRITE(5,1700)EB,SB,FNB
      WRITE(5,1700)EW,SW,FNW
      WRITE(3,470)IDECK0
      WRITE(3,1700)EB,SB,FNB
      WRITE(3,1700)EW,SW,FNW
1700  FORMAT(3F)
      GO TO 450
1800  CONTINUE
      END
```

THE EFFECT OF GEOGRAPHICAL DISPERSION OF THE COLLECTION ON DOCUMENT RETRIEVAL

T. Bartelt, F. Mundt, C. Wanat

ABSTRACT

Document retrieval, as part of an interlibrary loan operation, at the University of Illinois is discussed. A flowchart of the document retrieval procedure is used to illustrate the details of the process. Results of a small experiment are shown to indicate that batching of requests to individual libraries in sizes ≥ 8 effectively eliminate geographical distance as an important variable. However, this savings must be traded off against the time delay in cumulating such batch sizes. A larger data collection effort might resolve this tradeoff and is advocated.

INTRODUCTION

There is a large body of literature within library science concerning state-wide interlibrary loan policies, procedures and practices ranging from analysis of TWX systems [1], through staff and/or system organizational schemes [2], to delivery systems [3]. This study covers a very specific aspect of the interlibrary loan process: the effect of dispersion of the collection on document retrieval time of the Illinois Library and Information Network as it operates in one of its four centers, the University of Illinois at Urbana-Champaign.

Some background concerning the University of Illinois phase of the network may aid in understanding the study. Interlibrary loan requests arrive via teletype or mail in the Interlibrary Loan Office, Room 405 of the library, where a staff member notes the search designators, i.e. call number or author, title, etc. and the libraries where the item might be found. The requests are then sorted into piles by library which are picked up by the "runner" for those libraries which he will visit on that trip.

Assignment of libraries to be searched is done by a staff member who uses experience and intuition as well as the availability of runners in determining which libraries will be visited on which days and in which groupings.

Upon returning to the 405 office, the "found" items are placed on the appropriate shelving for shipment to the requesting agency. Records are kept on all materials sent out. However, shipment is not performed at the 405 office per se, that is handled by the University mailroom personnel or by the shuttle drivers from the Chicago Suburban Library Systems. Articles within periodicals are verified by a staff member and then sent to be photo copied. The articles themselves are not sent out but are returned to the lending library from which they were obtained. Only the photo copy goes to the requesting agency

A staff member also screens the returned requests separating those which are not available (so the requesting agency can be informed via teletype) from those which require a search-followup, special packages, etc.

SEARCH PROCEDURES

As the dispersion of the libraries is most noticeable in the effect it has upon the runner's operation when he picks up the requested items, it is this area in which our study was directed.

We discovered that there was no officially recommended procedure for the runners to follow. The first part of our study therefore was a study of the methods used by the runners to determine those factors which might affect the retrieval of requested materials.

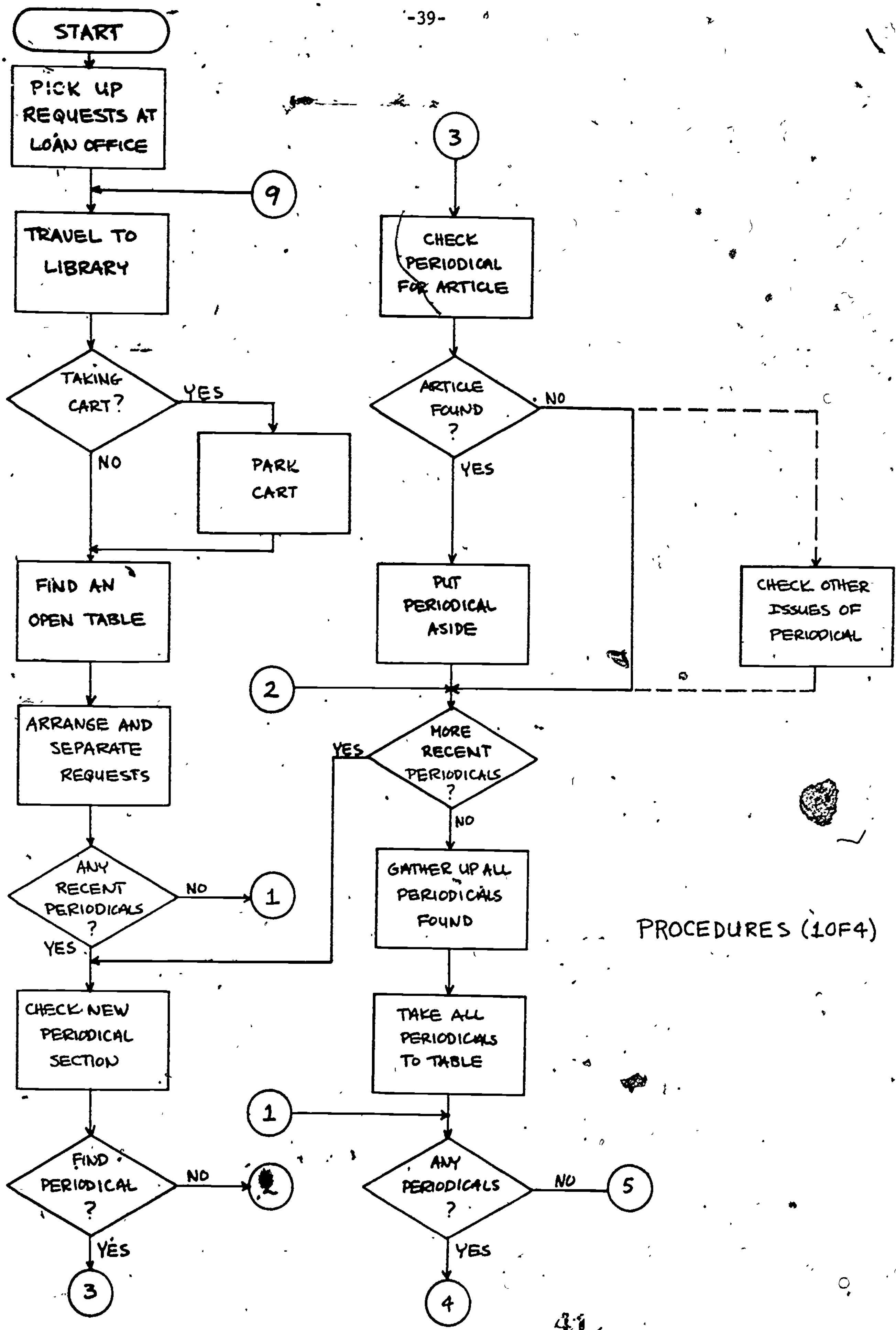
As a result, we discovered that a relatively uniform procedure was followed in all libraries except the Chemistry Library which requires all periodicals to be copied in the library. It was also noted that the Law Library also was atypical in that it employed no standard classification system.

A byproduct of this was the following flow chart which gives the steps followed by a runner in one trip.

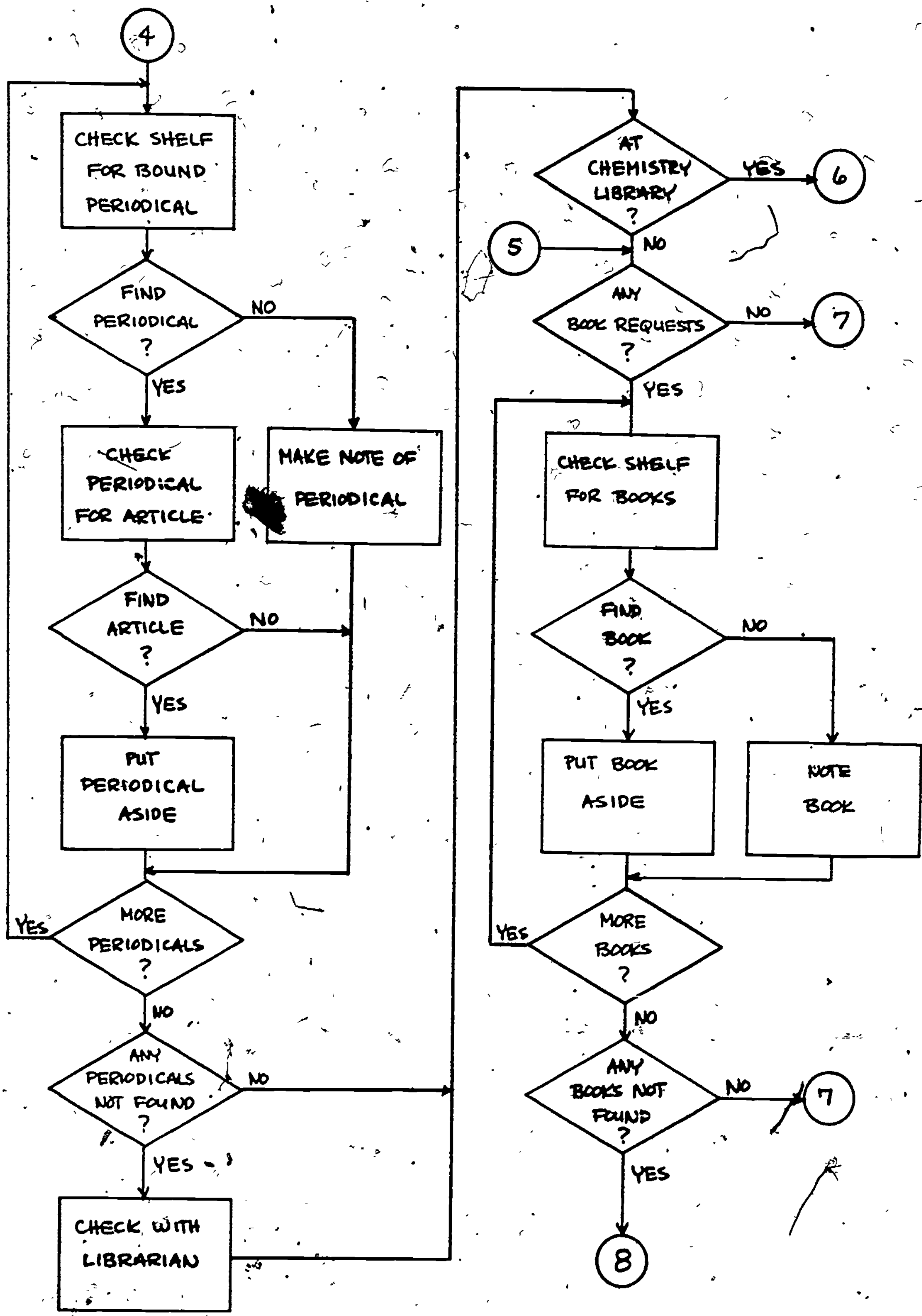
DATA COLLECTION

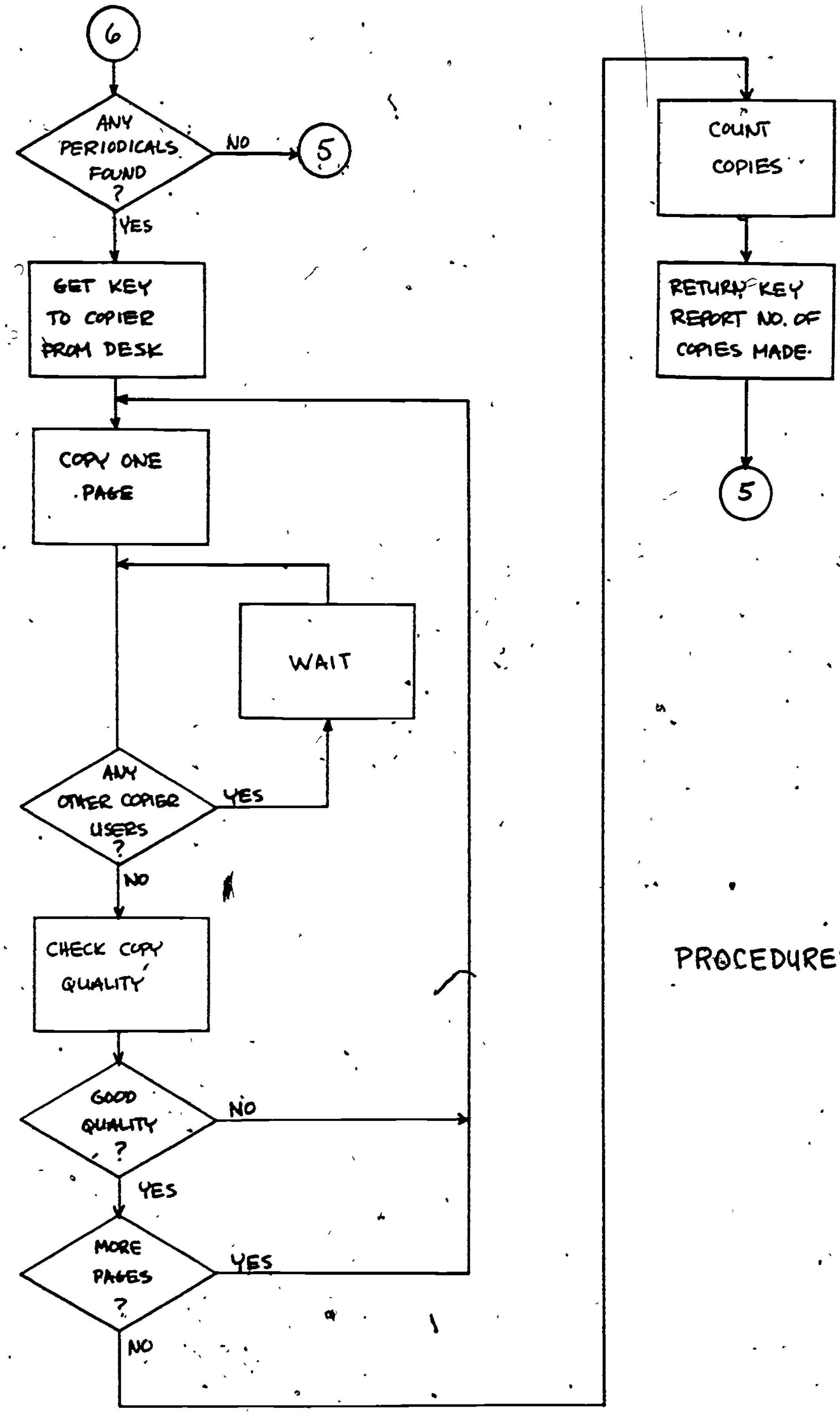
Since the point of the study was to consider the effects that decentralization (both in terms of distance and of administration) has on retrieval time, the data collection centered on time. A data worksheet was used which asked the runner to note the time he left the Interlibrary Loan Office, the time he entered a particular library, the time he left that library and the time he arrived back at the office. Also, within a library, the runner was asked to note how many requests he took to the library and how many he was able to fill. Since a runner used one sheet for each trip out of the office (and not for each library visited), it was also possible to note which libraries tended to be grouped together. A sample worksheet is appended.

Since runners would usually go to more than one library on any given trip, it was difficult to arrive at a travelling time per library. Eventually it was decided to use as this travelling time the total time spent en route for the trip divided by the number of libraries visited on that trip. While this is satisfactory in most instances (since the libraries grouped together tended also to be similar in terms of location with respect to the office), it certainly gave misleading times in some cases. For example, when a library close to the office was visited on the way to a distant one and the travelling time was the total divided by two, both of the individual travelling times are off considerably. (If a controlled experiment could be conducted with runs to individual libraries, perhaps this problem could be solved). With the data we had

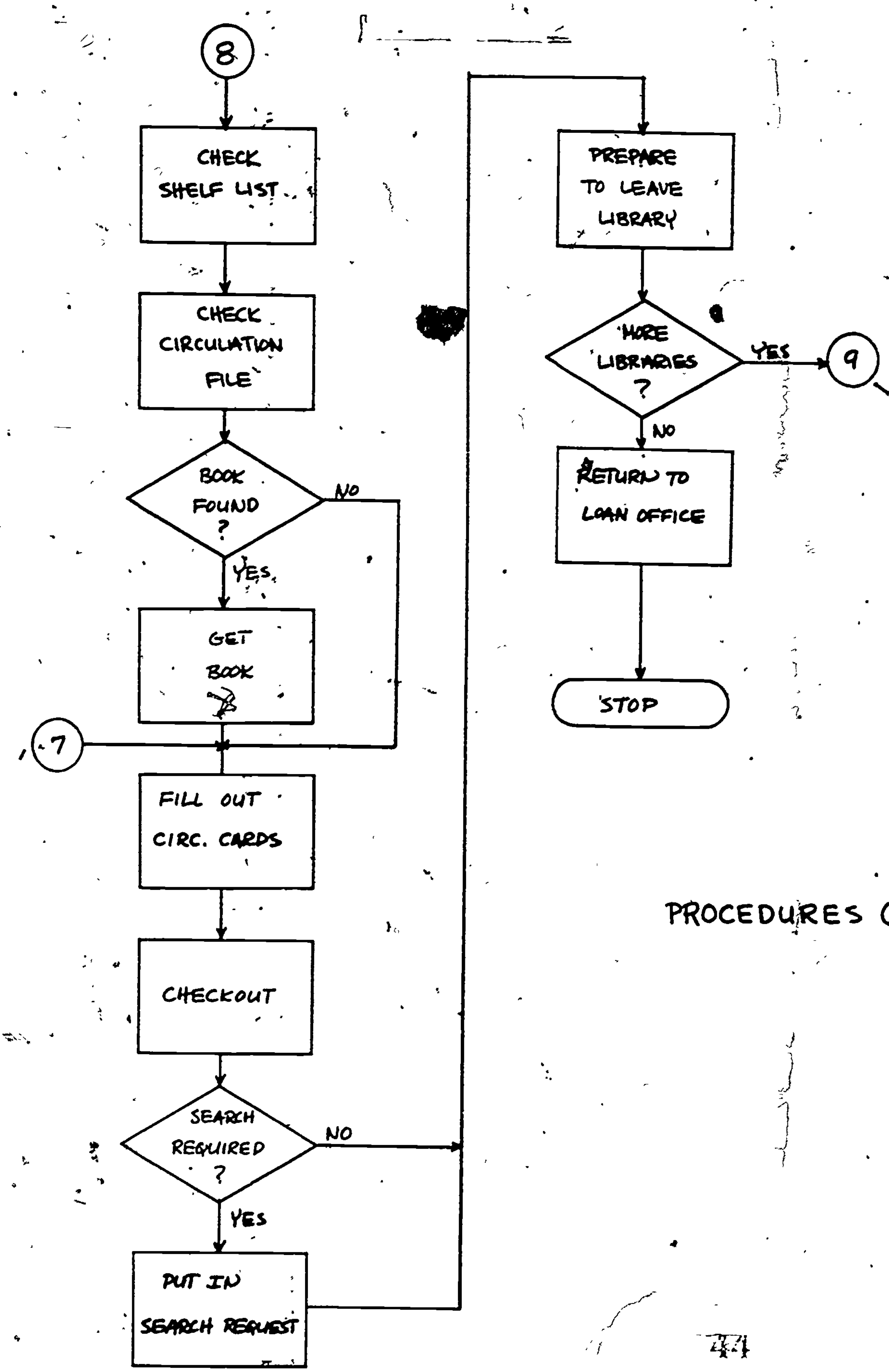


PROCEDURES (1OF4)


4
CHECK SHELF FOR BOUND PERIODICAL
FIND PERIODICAL ?
NO
YES
CHECK PERIODICAL FOR ARTICLE
MAKE NOTE OF PERIODICAL
FIND ARTICLE ?
NO
YES
PUT PERIODICAL ASIDE
MORE PERIODICALS ?
YES
NO
ANY PERIODICALS NOT FOUND ?
NO
YES
CHECK WITH LIBRARIAN
AT CHEMISTRY LIBRARY ?
YES
6
NO
5
ANY BOOK REQUESTS ?
NO
7
YES
CHECK SHELF FOR BOOKS
FIND BOOK ?
NO
YES
PUT BOOK ASIDE
NOTE BOOK
MORE BOOKS ?
YES
NO
ANY BOOKS NOT FOUND ?
NO
7
YES
8



PROCEDURES (3OF 4)



PROCEDURES (4 OF 4)

available, error could not be avoided and this approach seemed reasonable, if not exactly ideal.

A total time per library visit was determined as the sum of the travelling time plus the time spent in the library. Dividing this by the number of requests taken to the library gave us the time per attempt.

Due to shortage of time and low return rate of the forms, only 33 data points were obtained from 16 libraries; 9 of the libraries being within the main building and 7 outside it.

A comparison of distance versus average time for each item requested resulted in Table I where libraries 1 through 9 are within the main library building and 10 through 16 are outside in order of increasing distance from the main building.

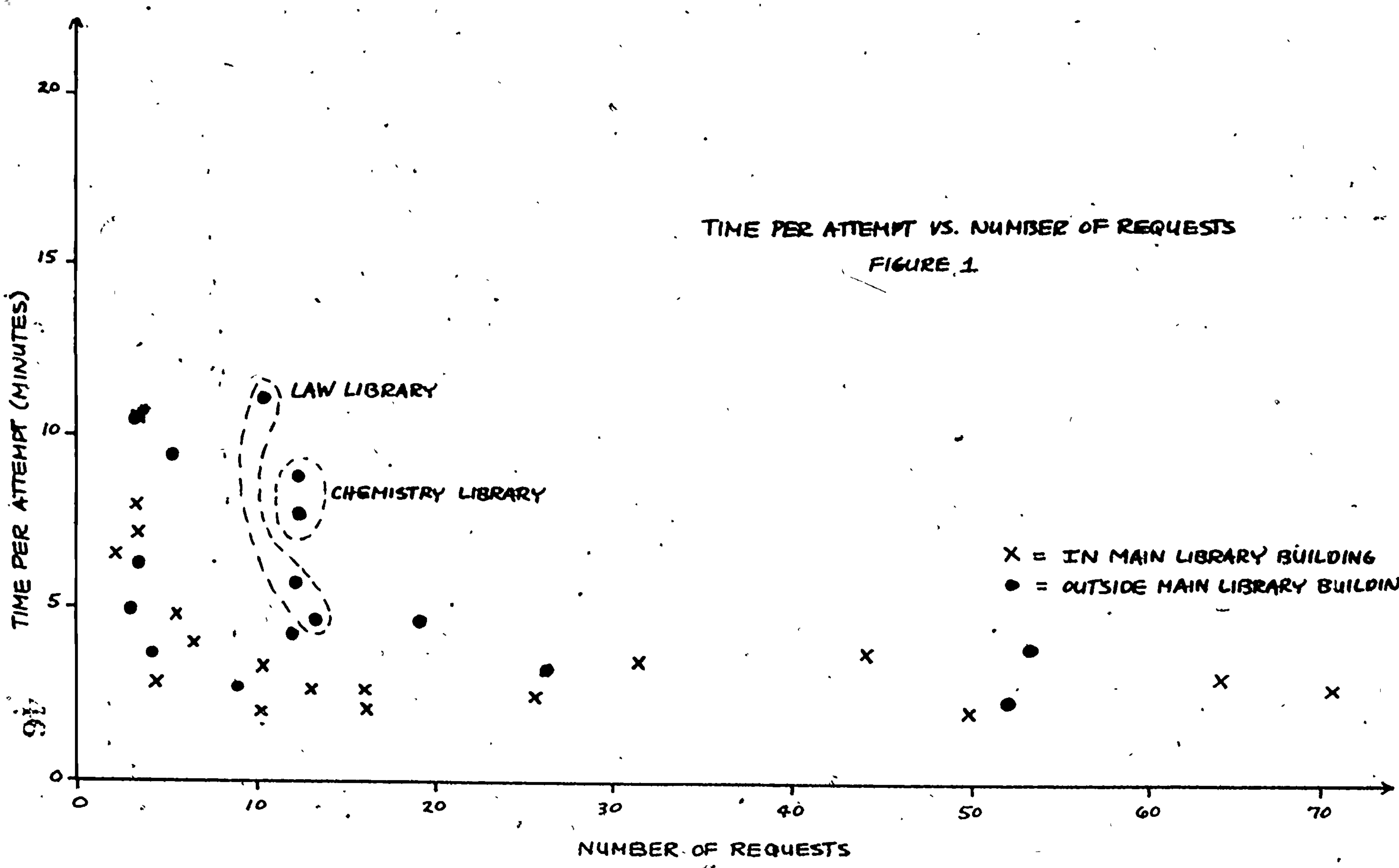
Library	Average Time (Minutes) Per Item Searched
1. Stacks	2.8
2. Reference	7.5
3. Education	2.9
4. English	4.7
5. Commerce	2.8
6. Illinois Historical Society	2.8
7. Library Science	4.8
8. Physical Education	3.4
9. Special Languages	2.2
10. Undergraduate	3.7
11. Communications	10.7
12. Home Economics	7.3
13. Natural History Survey	6.7
14. Law	7.1
15. Music	2.8
16. Chemistry	8.3

Table I

This would seem to indicate that those requests not in the building take almost twice as long to fill as those in the building. However, a closer look at the data results in Figure 1, graphing number of requests versus time per attempt. (Chemistry and Law libraries specified due to their special problems as previously noted). Not much difference is noticed between those libraries in and out of the building except, when libraries out of the main building have smaller numbers of requests, longer times result.

TIME PER ATTEMPT VS. NUMBER OF REQUESTS

FIGURE 1



CONCLUSIONS AND RECOMMENDATIONS

It would appear that the dispersion of the collection at the University of Illinois does result in an increase in document retrieval time, but that this effect is minimal if no library is visited with less than 8 requests. Of course, in waiting for requests to accumulate to batch sizes of 8 or greater, requests would experience additional delay. If arrival rate data were collected for items for each library, we could predict the average additional delay due to waiting for appropriate batch sizes. If this exceeded the time saved in retrieval, then such batching would be inappropriate. As a by-product, such batching might allow staffing decreases and/or increased productivity.

For batch sizes smaller than 8, increased retrieval time is due to both distance and within library processing as is shown by the increased time per request for small batches retrieved within the main building. Thus, distance is not the only factor involved in retrieving small batches.

Our conclusions are very tentative due to the lack of data caused by a period of system breakdown while the system was under study. (The teletype broke down giving an atypical period during which the data was ignored and not included in this study.) It is highly recommended that further use of the data gathering form be made and the data kept and made available for further study.

It is also suggested that the operation of the runners be standardized with some instructions, possibly using the enclosed flow chart if it is found satisfactory for this purpose.

ACKNOWLEDGEMENTS

Special acknowledgement and thanks for help in understanding the overall interlibrary loan operations; examining the problem's attack; becoming acquainted with the component parts and help in gathering data used in this study is given to the following full-time University of Illinois Interlibrary Loan staff members: Elaine Albright (Director of the Interlibrary Loan Office), Patricia Hausman, Rachmimi Loh, Stephanie Sebock, Carol Verniglea and Kitty Watters and to the following students who work under Ms. Sebock and on whom so much of the essential daily operations of interlibrary loan depends: Michael Barr, Jan Grossberg, Brent Grossman, Connie Hansen, Karl Jokela, Sara Lindbeck, Deborah Owen, Edward Regan and William Stewart.

REFERENCES

The following references deal with interlibrary loan policies and procedures or with decentralization in general. They are not overly applicable to the present study but may be useful for providing general background information concerning these two areas.

1. Blande & Holt, "Cost-performance Analysis of TWX Mediated Interlibrary Loan in Medium Sized Libraries," Med. Lib. Assn. Ba. 59: 65-70 (1971)

2. Daggan, M., "Library Network Analysis and Planning," J. of Library Automation, 2:157-75 (1969)

3. Herner, Saul, Recommended Design for the U.S. Medical Library and Information System, Wash. D.C., Herner & Co., 1966.

4. B. C. Brooks, "The Viability of Branch Libraries," J. of Librarianship. 2: 14-21 (1970).

5. J. Raffel and R. Shishko, "Centralization vs. Decentralization: A Location Analysis Approach for Librarians," Special Libraries, 63: 135-43 (1972).

Date: Initials: Time leaving ILL: Time returned ILL:	Time - Entering Library	Time - Leaving Library	Number of Requests Attempted	Number of Requests Filled
Stacks				
Classics				
Commerce				
Education				
English				
History				
Library Science				
Map & Geography				
Modern Language				
Physical Education				
Agriculture				
Architecture				
Biology				
Chemistry				
City Planning				
Communications				
Engineering				
Geology				
Health Sciences				
Home Economics				
Labor & Industrial Relations				
Law				
Mathematics				
Music				
Natural History Survey				
Physics				
Undergrad				
Veterinary Medicine				