

Subject Searches Using Two Catalogs: A Comparative Evaluation

This paper reports the results of a study undertaken to determine if dividing a traditional dictionary catalog would result in an increase in the effective use of university library catalogs. Two catalogs—one in dictionary arrangement, the other divided into subject and non-subject entries—were selected and the appropriate sections matched. Participants were chosen at random from the undergraduate population of the two universities. The results indicated that, for a series of questions representing different levels of difficulty, a change in arrangement from dictionary to divided would not materially assist college undergraduates in finding subject references.

TO THE LIBRARY ADMINISTRATOR, the catalog represents a substantial investment in funds, time, and personnel. Still indispensable as an index to the collections of most libraries, the catalog is nevertheless criticized for its limitations as a wholly effective tool for librarian or patron. Faced with selecting the best method to prepare, arrange, and maintain the catalog, the administrator must weigh a complex combination of factors that represent two basic variables, cost and effectiveness. He can bring to his consideration a wealth of testimony but very few objective findings. This paper

reports the results of a study investigating the proposition that dividing the catalog will result in improved effectiveness for patrons seeking a subject approach to the library's collection.

The Divided Catalog

Although a catalog might be divided on a number of different bases, the term "divided catalog" is commonly understood to denote an arrangement whereby the subject entries and the author and title entries are put separately into two alphabetical sequences. This plan differs from the "dictionary catalog," which places all three sorts of entries into a single sequence. The concept of the divided catalog can be traced to an article written by William I. Fletcher, librarian of Amherst College in 1905. Concerned that the dictionary catalog could not continue to cope with the complexities arising from the ever-increasing size of library collections, Fletcher advocated removing the sub-

Dr. Krikelas is Assistant Professor of Library Science in the University of Wisconsin. This article is based on the author's paper, "The Effect of Arrangement on the Successful Use of Library Catalogs" (unpublished PhD dissertation, University of Illinois, 1967). The study was done at the Library Research Center, University of Illinois, under a grant from the U.S. Office of Education (OEG-3-7-070014-1630).

ject entries to a separate file as he had done in his own library.¹ His proposal evoked little response and thirty years passed before the divided catalog was again prominently espoused. In 1935 Donald Coney asserted that "the catalog confuses the user with a wealth of detail in unfamiliar form." He suggested that the dictionary catalog be simplified by dividing it.²

In the years following Coney's article, a substantial body of literature on the subject has been produced.³ A review of this literature reveals that no previous study has attempted to establish a clear relationship between the type of arrangement and the successful use of the catalog. Although writers since Coney have seen in the divided catalog a promise of simplified filing and reduced congestion at the catalog, these supposed benefits are elements of the cost variable rather than the effectiveness, which is the focus of this study. These articles and other specific investigations of catalog use, summarized by Stevens,⁴ Tauber,⁵ and Frarey,⁶ have identified a number of obstacles to effective searches of the catalog for subject references. These difficulties are relevant to the problem under investigation.

Occasionally the patron expects to lo-

cate material, such as periodical articles, that is not traditionally analyzed in the catalog. Sometimes he fails because he bases his search on incorrect bibliographic information. Perhaps most often he is unable to select the appropriate term, or he approaches the catalog at a different level of specificity than is necessary to achieve a successful search.

Dividing the catalog would not solve all of these problems. For the most part, these difficulties stem from a lack of sophisticated knowledge by the patron rather than from the arrangement of the catalog itself. Division, however, might well reduce the confusion between subject and non-subject entries. One example of such confusion has been reported by Margaret C. Brown. Observing a graduate student seeking information on the subject "Rural Recreation" she noted:

Next the student went in search of any subject which began with the word "rural." Here several titles relative to the subject were found. The student was highly pleased with this development but quite unaware that these were title entries.⁷

The number of potential conflicts is higher than at first might be expected. Various rules of cataloging tend to suppress similarities between the subject of a book and the title of that book in favor of the subject entry. Other subject headings, however, are identical in form to main or non-subject added entries. Hence references to material by or about an individual, society, institution or governmental agency will use the same terms for the heading whether they are main or secondary entries. To emphasize the difference between identical headings for different concepts, two general devices are used. One method is to vary the typographical presentation

¹ William I. Fletcher, "The Future of the Catalog," *Library Journal*, XXX (March 1905), 141-44.

² Donald Coney, "The Librarian and the Catalog," *ALA Bulletin*, XXIX (September 1935), 593-94.

³ Bibliographies on the subject include Julia Pettee, *Subject Headings* (New York: H. W. Wilson Co., 1947), p. 185-86; Dorothy Grosser, "The Divided Catalog: A Summary of the Literature," *Library Resources & Technical Services*, II (Autumn 1958), 238-52; Maurice F. Tauber, "Cataloging and Classification," Ralph R. Shaw, ed., *The State of the Library Art*, I, Part 1 (New Brunswick, N.J.: Graduate School of Library Service, Rutgers—The State University, 1960), p. 92-101; and Theodore C. Hines and Jessica L. Harris, *Computer Filing of Index, Bibliographic and Catalog Entries* (Newark, N.J.: Bro-Dart Foundation, 1966), p. 105-106.

⁴ Rolland E. Stevens, *A Summary of the Literature on the Use Made by the Research Worker of the University Library Catalog* (University of Illinois Library School, *Occasional Papers*, No. 13, Urbana, 1950).

⁵ Tauber, *op. cit.*, p. 65-101.

⁶ Carlyle J. Frarey, "Subject Headings," in Shaw, *op. cit.*, I, Part 2, p. 49-50.

⁷ Margaret C. Brown, "The Graduate Student's Use of the Subject Catalog," *College & Research Libraries*, VIII (July 1947), 203-08.

by indicating subject entries in red ink or in capital letters and all other entries in conventional upper and lowercase form. The second device is to treat each type of heading as a separate file. In the dictionary catalog, subject cards are filed after the identical headings for main and added entries; in the divided catalog, all subject cards are removed as a body to a physically separate file.

Some evidence and much testimony indicate that the typographical devices are often too subtle for the lay user of the catalog to recognize.⁸ In large or highly specialized catalogs the great number of possible conflicts compounds the confusion. Since the divided catalog, at least in theory, tends to identify the subject entries unambiguously by segregating them, there is reason to suppose that the divided catalog would be more effective for subject searches than the dictionary catalog. Therefore, if the same person made identical subject searches in two catalogs, one divided and one dictionary, it can be assumed that his difficulties would be common to both catalogs except that, in the divided catalog, title and other conflicting entries would not be confused with subject entries. This concept can be restated into the following specific hypothesis and tested empirically:

Assuming all other factors are equal, subject searches using a catalog in which the subject entries have been separated (*i.e.*, a divided catalog) will produce more pertinent references and fewer inappropriate references than identical searches using a file combining all entries into a single (dictionary) sequence.

Design of the Study

Ideally, the simplest design for testing an hypothesis of this nature would be to have the same patrons conduct the same searches twice, first with a cat-

alog of a given arrangement and, second, with the same catalog after it had been rearranged in ways assumed to improve it. A comparison of the amount of success achieved using each form of the catalog would measure the effect of the modification of the catalog on success in the searches. In practice, however, it is difficult to control the possible effect of a particular patron's prior experience and familiarity with one of the two catalog arrangements. Moreover, the second search is likely to be biased toward success by the patron's memory of the first search. Moreover, modifying the catalog to set up "before-after" comparisons would seriously inconvenience an operating library and its patrons.

An alternative design is to locate two catalogs that are largely alike except in arrangement. By judicious control of the type and number of searches to be made, the study can be restricted to specific sections rather than the whole of the catalogs. Use of catalogs at different institutions, however, makes it difficult to employ the same patrons as participants in both parts of the study.

An accepted solution to this problem of design is to match individuals at the two institutions, to give them identical search problems, and to treat the results as those of one person. Matching individuals presents obvious difficulties, but as is pointed out in one research text:

The more precise the matching and the greater the number of factors on which matching is to take place, the greater the number of cases for which no match is available. Fortunately, however, relevant factors are often so interrelated that matching on one factor brings with it partial matching on other factors; there is a "diminishing return" as additional factors are controlled.⁹

The key problem then is to select for purposes of matching those characteris-

⁸ See for example: Earl Farley, "Rubrication: A Special Library Art Transforms a General University Catalog," *Special Libraries*, LIII (July-August 1962), 330-31.

⁹ Claire Selltiz, *et al.*, *Research Methods in Social Relations* (Rev. ed.; New York: Holt, Rinehart and Winston, 1959), p. 105.

tics that are most closely related to successful use of library catalogs. No previous study has succeeded in reaching definitive conclusions as to the characteristics that are clearly crucial to success in catalog use. It was necessary, therefore, to assume in this study that the two most important factors are experience and familiarity. For the user-universe of college undergraduate students selected for this study, these factors were measured by the student's class standing expressed in number of semesters on campus and the frequency of use of the main college catalog. Evidence as to the validity of this basic assumption was gathered in the course of the investigation, and the analysis of these data is reported as part of the results of the study.

Beyond selecting the catalogs to be used and the respondents to be tested, it was necessary to plan the pattern of the actual searches. Traditionally, catalog-use studies have observed an individual at the catalog, have noted the purpose of his search, and then have judged success or failure from a determination by the investigator or by the respondent himself that what he found did or did not achieve the original purpose. Under such circumstances the interpretation of success or failure itself can be questioned and, in any case, many other factors than the catalog alone are likely to be involved in the outcome.

In order to limit variations to differences in catalog arrangement alone, it was deemed advisable to develop in advance a battery of test searches rather than to leave their selection to the respondents. The problems were chosen to include examples of identical headings that represented main entries, subject entries, and other added entries. Further, to determine whether a confusion of type of entry is the actual source of difficulty rather than simply a

general perplexity, the problems required use of a selection of traditional subject headings.

Evaluation of the success or failure of the searches was planned to be as objective as possible. Degree of user satisfaction was rejected from the outset as a sufficient test of effectiveness. Even reliance upon the fact of locating an apparently relevant document was not considered enough, since relevance is a function of the subjective purpose of the user. Therefore, it was decided in advance to make the critical test the location of cards bearing predetermined subject headings rather than references to any specific documents.

A card was determined to be relevant only if it contained the exact subject heading requested. It was anticipated that participants would indicate all, some, or none of the pertinent references and none or some non-pertinent cards. In order to make the results comparable a "success ratio" for each search was computed. By taking an average of these success ratios for each student, a "mean success" score was determined for each participant. In the development of these scores three factors were considered: (1) the number of relevant references retrieved; (2) the predetermined total number of relevant references in the file; (3) the total number of references retrieved (whether relevant or not) by the participant. If the number of relevant references retrieved was zero, the success ratio, by definition, was scored as zero.

It must be emphasized that this success score was developed only to identify the relative success of the individual participants in locating pertinent references. In no way was it proposed that this measure would also be appropriate for other types of tests. Furthermore the success ratio did not indicate why some participants had only partial, rather than complete, success. Such analysis of

causes of failure was undertaken separately and is also reported. For a given set of structured search-problems, however, the users of the two types of catalogs could be compared meaningfully and the expected differences between the matched pairs tested for statistical significance.

Collection of the Data

Because of certain basic similarities, two large universities were selected as the setting for this study—one with a conventional dictionary catalog and the other with a divided (author-title and subject) catalog. Both schools are large, midwest, state-supported institutions of national reputation. Although not identical, the two schools are also similar in the distribution of the undergraduate student body according to class standing. On the other hand, there is a discrepancy in the size of the two collections that could have an effect on the attempt to locate comparable sections in the two catalogs. A preliminary inspection of both catalogs, however, indicated that many of the search-problems selected at one school involved use of file sections similar in size at both schools. The catalogs were also alike in many other ways with only one major difference deemed to be a potential difficulty.

The one important difference occurred in the matter of filing. In particular, the rule affected the filing order of subordinate agencies and had direct applicability to the study. In the divided catalog, subordinate agencies (*e.g.*, The United States Civil Service Commission) were filed after the general heading "U.S." and its appropriate subject subdivisions. In the dictionary catalog, these subordinate agencies were considered, for filing purposes, to be indistinguishable from subject subdivisions. For example, subject entries for various headings under "United States" appear in the two catalogs in the following order:

Divided Catalog

U.S.—ALTITUDES
U.S.—BIBLIOGRAPHY
U.S.—CIVILIZATION
U.S.—FOREIGN RELATIONS
U.S.—HISTORY
U.S.—POLITICS & GOVERNMENT
U.S.—STATISTICS, VITAL
U.S.—TERRITORIAL EXPANSION
U.S. ARMY
U.S. ARMY—BIBLIOGRAPHY
U.S. LIBRARY OF CONGRESS
U.S. WEATHER BUREAU

Dictionary Catalog

U.S.—ALTITUDES
U.S. ARMY
U.S. ARMY—BIBLIOGRAPHY
U.S.—BIBLIOGRAPHY
U.S.—CIVILIZATION
U.S.—FOREIGN RELATIONS
U.S.—HISTORY
U.S. LIBRARY OF CONGRESS
U.S.—POLITICS & GOVERNMENT
U.S.—STATISTICS, VITAL
U.S.—TERRITORIAL EXPANSION
U.S. WEATHER BUREAU

Thus a search for subject cards for the entry "U.S. Civil Service Commission" should be made between "U.S. Army—Bibliography" and "U.S. Library of Congress" in the divided catalog. In the dictionary catalog, "U.S. Civil Service Commission" would appear between the entries for "U.S.—Bibliography" and "U.S.—Civilization." Rather than eliminating such examples, it was decided to include them and, through analysis, to determine if this was the cause of success or failure more frequently at one school than at the other.

The actual selection of the search problems was achieved by random sampling from the dictionary catalog. The objective of the sampling was to obtain a list of personal, corporate, and uniform entries that could be compared with the divided catalog. A list of two hundred conventional subject headings was also compiled by sampling from the Library of Congress subject headings.¹⁰ The dictionary catalog was re-audited to determine if these subject headings

¹⁰ *Subject Headings Used in the Dictionary Catalog of the Library of Congress* (6th ed.; Washington: Library of Congress, 1957).

were actually used and to assess the feasibility of including them as problems.

Prior to the final comparison of the appropriate sections of the two catalogs, a pre-test was conducted. The purpose of the pre-test was: (1) to determine the effect, if any, of the alternative ways of wording the questions; (2) to determine the total number of questions that might be tested in a one-hour period; and (3) to determine if patterns of search actually were similar enough to predict the general sections of the catalog that should be compared.

Analysis of the results of the pre-test indicated that the students' responses were related to the nature of the question rather than to the way the question was worded. It was also determined that the final exercise would require, on the average, about forty-five minutes. The remainder of the hour was set aside for explaining the procedure and for post-test interviews.

Observation of the procedure and patterns of search during the pre-test also supported the expectation that for specific requests the appropriate sections of the catalog could be determined. There were instances when the participant elected to search the catalog for less specific subjects than the ones requested in the problem. Interviews following the exercises revealed that rewording the subject request would in no way have made it clearer to the student that such a subject actually was used in the catalog. Of particular interest was the relatively consistent pattern of search whether it was for the precise subject heading or for a less specific term.

The final comparison of the two catalogs was conducted two weeks prior to the commencement of the testing. To test the assumption that known-item searches would not be affected by the arrangement of the catalog, two ques-

tions asked the respondent simply to locate given author and title entries. The remaining questions were subject searches and consisted of requests for cards concerning a specific work of one author (a literary criticism), three conventional subjects, two personal-name entries as subjects, and three corporate entries as subjects.

The participants were selected at random from the undergraduate population of the two universities. Since lists of the student population by classes could not be secured, the student directories of the respective schools were used. Random selection was used only to reduce any unknown bias in response rate that might have resulted from other selection techniques. The anticipated statistical tests also indicated the desirability of having at least thirty matched pairs. Letters requesting the participation of the students were mailed so as to arrive during the first day of the second semester of the 1966-67 academic year.

In all, 171 students took part in the study although twenty-three were unable to complete the search-problem exercises within the allotted period of time. In addition, four students listed as underclassmen in their respective directories were found to be enrolled in professional programs and in their fifth or sixth year of college. These students were also deleted from the study. The final number of usable scores was fifty students using the divided catalog and ninety-four using the dictionary catalog, a total of 144 participants.

The procedure in conducting the search exercises was similar at the two schools. Students completed a "general information form" containing requests for personal data before proceeding. As each participant searched the catalog to find the appropriate cards, the investigator noted the procedure as well as the final decision. Every effort was made to

secure information about where the individual searched, what specific heading he had in mind, and the type of difficulty, if any, that he encountered. During the instruction period students were told that the test questions had been selected at random and that some of the requests might not represent areas of interest to them. It was assumed, however, that the student could cope with the request at a level determined by his basic knowledge about the catalog. At the end of each session, a post-test interview was conducted. One of the questions asked of the participant was "Did you find that the problems and your responses were a fair indicator of your general knowledge of the scope and arrangement of the catalog?" The replies of the students, admittedly testimonial, gave no reason to suspect the validity of this approach to measuring effective use.

Matching the Students

In very general terms, the matching was successful to the extent that thirty-one pairings were made. Although there were some minor differences between the matched pairs, every effort was made to have the two primary criteria ("semesters on campus" and "frequency of use of the main catalog") as equal as possible. Basic information about the participants was entered on cards and used to separate students into groups according to the number of semesters on campus. These groups were subdivided into categories according to the frequency of use of the main catalog, and subsequent matching was accomplished by scanning other characteristics.

In order to compare the two primary criteria with other potentially useful characteristics for matching, a series of questions was asked of the participants on a "general information form." Data on the following characteristics were

collected: (1) semesters on campus and class standing (also converted to semesters in college) as measures of exposure to the catalog; (2) frequency of use of the main catalog as a measure of familiarity; (3) sex; (4) cumulative grade point average; (5) most common approach to using the catalog; (6) the type and amount of instruction received in "how to use the library"; and (7) work experience in libraries.

The procedure for analyzing these data was to investigate the relationship between personal characteristics and the student's performance. This analysis was based on the concept that one or more characteristics could be shown to be related to the mean success score at each school. By examining these characteristics it could be determined if they were common to the students at both schools and therefore generally applicable as criteria for matching. The analysis was made for each group of students so that for each school identical questions (a constant) were searched using the same catalog (also a constant) by different participants. Hence, the differences in success scores for each group could be attributed to the participant's conduct of the search, rather than to the question or the catalog.

The relationship between the various characteristics and performance was determined by computing some measure of association where applicable. For data given in interval measures, a Pearson product-moment correlation was calculated and the resulting correlation coefficients tested for significance using a table of expected values.¹¹ For data that lent themselves to natural dichotomies (*e.g.*, sex, instruction, and work experience), a point biserial correlation coefficient was calculated and tested for

¹¹ Allen L. Edwards, *Statistical Methods for Behavioral Sciences* (New York: Holt, Rinehart and Winston, 1954), p. 142-55.

statistical significance using a *t*-test.¹² Finally, for a few of the characteristics, mean success scores were grouped and the differences between the means of these groups were tested using analysis of variance.¹³

It should be noted, however, that these analyses (only summarized here) must be considered tentative at best. The purpose of the random selection procedure was to minimize response bias and to ensure the best chances for matching. The sample cannot be considered—nor was it intended—to be a true probability sample of the undergraduate population at either school. Therefore the inferences apply to the participants only.

The analysis of the relationship between personal characteristics and mean success score for the study groups indicated no significant association except for grade point average. This relationship was limited to the larger study group using the dictionary catalog and even here the relationship was so low that it accounted for less than 5 per cent of the variance. Nonetheless the matching procedure was reviewed in respect to grade point average, but there was no evidence that the matching procedure followed was in any way invalid. In fact, given a much larger sample from each institution the simple process of random pairing without attention to these characteristics would have been appropriate.

An effort to determine the fairness, if not the validity, of the test exercises was also undertaken. Although the mean score of each question was expected to vary from school to school, it seemed reasonable to expect that the relative difficulty encountered in the eight prob-

lems would be the same for each group of students. Therefore, it was predicted that if each question were ordered by degree of difficulty as represented by the mean scores, the rank order would be the same for both groups. A Spearman-rho rank order correlation coefficient (r_s) was computed, and since the calculated value, $r_s = .922$, was significant at the .05 level, it was concluded that each question represented the same degree of difficulty and was not biased in favor of either group.¹⁴

Analysis and Results

The primary objective of the study was to test the hypothesis that dividing the library's catalog would permit improved use of the catalog by library patrons. For the specific empirical test described, the original hypothesis can be restated as follows:

Assuming all other factors are equal, the mean success score for an individual using a divided catalog will be significantly greater (statistically) than the resulting score for the same searches using a dictionary catalog.

The test for significance is one of computing a *t*-statistic by dividing the difference between the means for each group by the standard error of the difference for the matched groups. Mathematically this would appear as:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_{\bar{X}_1 - \bar{X}_2}}$$

For thirty-one matched pairs (*i.e.*, thirty degrees of freedom), the expected value of the *t*-statistics at the .05 level for a one-tailed test is 1.697. That is, a value of *t* calculated from the test group can be expected to be 1.697, or less, by

¹² Helen M. Walker and Joseph Lev, *Statistical Inference* (New York: Henry Holt and Company, 1953), p. 262-67.

¹³ William L. Hays, *Statistics for Psychologists* (New York: Holt, Rinehart and Winston, 1963), p. 356-458.

¹⁴ Sidney Siegel, *Nonparametric Statistics for the Behavioral Sciences* (New York: McGraw-Hill Book Company, Inc., 1956), p. 202-13.

chance alone ninety-five out of one hundred times.

The experimental data were tested and found not to be significant. Specifically, the value of the experimentally derived t was:

$$t = \frac{.456 - .393}{.0411} = 1.537.$$

From this evidence, there is no reason to reject the null hypothesis that no difference exists between the two groups. Since subsequent analysis showed that one of the questions was affected by some intervening factor, the test was recomputed for only eight subject searches. Again, while the value of the calculated t -statistic was higher (1.678), it too was not significant at the .05 level. Assuming the validity of the various underlying assumptions appropriate to the procedure for matching and testing, the divided catalog did not appear to be more effective for the participants.

In order to collect evidence to support the hypothesis that differences in arrangement do not affect known-item searches, two non-subject requests were included in the exercises. Table 1 is a summary of the per cent of success and failure in locating the call number for a book by Ernest Nagel. The table is based on the responses of all students completing this problem.

TABLE 1
RESULTS OF THE SEARCH FOR THE
CALL NUMBER FOR THE "NAGEL" BOOK

| Catalog Used | Number of Students | Per Cent Found | Per Cent Not Found |
|--------------|--------------------|----------------|--------------------|
| Divided | 51 | 98 | 2 |
| Dictionary | 103 | 98 | 2 |

The second question requested the reporting of the call number for the Warren Commission Report. The responses for all students completing the search are given in Table 2.

TABLE 2
RESULTS OF THE SEARCH
FOR THE CALL NUMBER FOR THE
"WARREN COMMISSION REPORT" QUESTION

| Catalog Used | Number of Students | Per Cent Found | Per Cent Not Found |
|--------------|--------------------|----------------|--------------------|
| Divided | 54 | 61 | 39 |
| Dictionary | 104 | 67 | 33 |

An analysis of the data (chi-square test) indicated that the small differences between the two groups were no more than might be expected by chance alone. It was concluded therefore that arrangement had no effect on known-item searching.

During the data collection process it was observed that a number of factors could be considered as the possible causes for failure or partial success. The three major reasons were (1) the use of incorrect search terms; (2) difficulty with filing rules; and (3) the inability of the patron to distinguish subject entries from non-subject entries. In addition, a number of participants also were affected by small peculiarities in the catalog (common to both schools) or other small problems particular to the individual.

The use of incorrect search terms. The largest single cause associated with complete failure in locating appropriate cards for any request was the selection of incorrect search terms. In an attempt to minimize this difficulty, the questions submitted to the students were in terms that appeared in the catalog. In some cases, these were unused terms for which a cross-reference was available. Nevertheless, of the 1,152 searches included in this analysis, 334 (29 per cent) were conducted using the wrong term or terms.

For most questions, the headings selected were more general than the request itself. For example, material about (rather than by) the Amateur Athletic Union of the U.S. was sought under the

more general heading "athletics" rather than the specific name of the organization. This approach was commonly used for other corporate headings. The only type of search that did not seem to pose a problem concerned requests for cards about individuals. Difficulty in selecting the correct search term for topical subjects varied with the nature of the heading. Fewer students had difficulty finding straightforward headings such as "Statistical Design," for example, than phrases such as "Chemistry as a profession." For the latter question, most participants searched under "Chemistry—Profession" and "Profession—Chemistry."

Difficulty with filing rules. Responses to two of the nine questions revealed that success was affected by various complexities of the filing rules used in library catalogs. It was anticipated that searches for which the appropriate entry is a subordinate governmental agency might be affected by the differences in filing rules at the two schools. To test the effect of filing, the two questions that were directly affected were analyzed. The problems, as presented to the students, were stated as follows:

Locate the appropriate catalog cards that indicate the library contains material about (rather than by): the U.S. Civil Service Commission.

Locate the appropriate catalog cards that indicate the library contains material about (rather than by): The Great Britain Board of Trade.

The analysis was based on an examination of the procedure followed at the two catalogs. Frequencies for those who chose the correct search term were tabulated. The category "found" includes all students who located the term even if the student made some subsequent error in selection of cards. The results of this analysis are summarized for the two questions in Tables 3 and 4.

The calculated value of chi-square for the measures given in Table 3 is

TABLE 3

RESULTS OF THE SEARCHES FOR THE
"U.S. CIVIL SERVICE COMMISSION"
SECTION IN EACH CATALOG

| Catalog Used | Number of Students | Per Cent Found | Per Cent Not Found |
|--------------|--------------------|----------------|--------------------|
| Divided | 15 | 53 | 47 |
| Dictionary | 51 | 98 | 2 |

21.72, which is significant at the .05 level for one degree of freedom. The result indicates that there was a significant difference in the performance of the students depending on the catalog (and the particular rule for filing).

TABLE 4

RESULTS OF THE SEARCHES FOR THE
"GREAT BRITAIN BOARD OF TRADE"
SECTION IN EACH CATALOG

| Catalog Used | Number of Students | Per Cent Found | Per Cent Not Found |
|--------------|--------------------|----------------|--------------------|
| Divided | 31 | 67.7 | 32.3 |
| Dictionary | 84 | 60.7 | 39.3 |

The calculated value for the measures given in Table 4 is 1.02 however, which is not significant at the .05 level. This would indicate that the performance of the two groups was virtually identical and not affected by the filing differences. The conflicting results raise the question whether the filing rules actually affected the searches or whether other factors were present. It seems plausible to speculate that students with previous experience have become conditioned to the apparent difficulty of using the "U.S." files, but do not carry over their experience when using headings for other countries. This explanation, however, does not account for the differences between schools. Since there was a possibility that this single question concerning the U.S. Civil Service Commission might have biased the results, the question was deleted from the calculated mean for each student. The resulting comparison of matched pairs for

eight searches also supported the original conclusion that there was no difference in the results for subject searches between the users of the two catalogs.

Other problems. A number of problems varied from question to question and from individual to individual. These were relatively few in number—such as searching under “amateur”—and did not affect the overall comparison. One particular problem involving the use of a “see” reference, however, brought to light interesting factors. The question directed searchers to the heading “Statistical Design” where a cross-reference to the used term “Experimental Design” was found. This “see” reference was interpreted as having the same meaning as a “see also” reference by twenty-four (24.4 per cent) users of the dictionary catalog and by two (4 per cent) of the divided catalog users. A number of the dictionary catalog users indicated that the two title added entries appearing after the cross-reference were the only appropriate references and preferred to consider any entries under “Experimental Design” (which were not searched) as a last alternative. The two divided catalog users simply decided that nothing on the subject was available.

Perhaps more informative was the procedure of many of the students who searched the complete “Experimental Design” file but indicated, as pertinent, only those cards that had the term “statistics” in the title. During the post-test interview these students indicated that normally they would undertake such a search if the two title entries under “Statistical Design” had not proven useful. There was little question that these twenty-six students (at both schools) did not consider the two terms as synonymous but rather as a reference from a specific term to a more general, inclusive term.

Inability to distinguish between subject and non-subject entries. The major hypothesis of this study was directly

concerned with this category of difficulty. The fact that 23 per cent of the students had difficulty in distinguishing between the two types of entries makes this confusion the second largest contributor to failure or partial success in making subject searches. For the dictionary catalog searchers, this problem was primarily one of selecting a variety of non-subject added entries as being subject headings. The fact that such entries were not typed in red did not seem to matter.

For the divided catalog user, the results were even more enlightening. Of the four hundred searches tabulated for this analysis, ninety-one (22.75 per cent) used the author-title catalog for subject searches. Since the frequency of such searches varied with the question, it became evident during the study that the major factor was simply an accident of location. As the student considered the individual question, he would walk through the main catalog section and search for the appropriate alphabetical sequence. Whether this sequence was part of the subject section or the author-title section did not seem to make any difference. Nor did the students note that the two sections were clearly marked and that different colors were used for the drawer labels to distinguish the sections. Interestingly enough, all of these students had indicated that they had used this catalog at least once within the past semester. It is difficult to imagine what else the library staff could have done to make the division more explicit.

Conclusions

Every study has inherent limits—both conceptual and practical—that define the degree of generalization that is possible. In the interest of maintaining maximum control over the various elements of catalog searches, the choice of participants, catalogs, and questions was highly structured.

Effective use of the catalog was measured in terms of a mean success score for subject searches. This score represented the ability of students to select appropriate subject references in response to a series of questions. The experimentally derived data were tested for significance and found not to be different. It was concluded that dividing the catalog was not a satisfactory device for making subject searches more effective.

The analysis of difficulties students had in coping with various questions was undertaken to determine if such difficulties were associated with arrangement. The results of that analysis indicated that for any potential benefits attributable to the divided catalog (*i.e.*, a larger percentage of successful searches for one or more questions), there were corresponding disadvantages (in terms of lower rates of success for other questions). Furthermore, the per cent of failures attributed to the inability of patrons to distinguish subject headings from non-subject entries was almost as great for the users of the divided catalog (22.8 per cent) as for the users of

the dictionary catalog (23.4 per cent). It was concluded therefore that for the two groups in general, the divided catalog did not facilitate subject searches more than the dictionary catalog.

The effect of arrangement on known-item searches was also investigated. Students were asked to determine if their particular library contained two specific documents. An analysis of the responses indicated that the rate of success in obtaining the call number for the two documents was not related to the differences in arrangement of the two catalogs.

In summary, the results of this study indicated that, for a series of questions representing different levels of difficulty, a change in catalog arrangement would not materially assist college undergraduates in finding subject references. Neither arrangement proved to be substantially superior. The academic librarian choosing between a divided catalog or a dictionary catalog can base his decision on cost of production and maintenance with reasonable confidence that either arrangement is equally effective for undergraduates making subject searches. ■ ■

