

National Collecting Trends: Collection Analysis Methods and Findings

Anna H. Perrault
University of South Florida

The primary focus of collection evaluation and assessment has been the subjective judgment of the strengths and weaknesses of library collections. The process has generally been a local application, often utilizing peer group comparative data. Quantitative collection analysis to support the subjective processes of collection evaluation and assessment is now almost universally conducted through extraction of data from local systems or network databases. National collecting patterns can be studied through the use of data extracted from the bibliographic utilities. This article examines the use of data from the *OCLC/AMIGOS Collection Analysis CD* and standard statistical series as a methodology for studying national collecting patterns. Findings on national collecting patterns from 1986-1995 in academic libraries in the United States utilizing the methodology are presented.

In a paper assessing the current state of bibliographic control, Warren Haas, in 1977, used a striking phrase—“*if we employ managerial instinct for lack of research...*” He continued on with an example, “...we might *estimate* that the nearly 8 million volumes added each year to just the university collections included in the Association of Research Libraries membership involve a total of perhaps 300,000 discrete titles” (Haas 1977, p. 276). The eight million volumes is a figure which could be obtained from the Association of Research Libraries statistical series; the 300,000 discrete titles was a sheer guess—instinct rather than research. Haas was not advocating instinct over research; he was simply acknowledging that no data were available to help him make his point.

This article was presented at the ALA Library Research Round Table program, Washington, DC, June 28, 1998. The research was partially funded by a grant from the University of South Florida Research Council. The author would like to express appreciation to the administration and staff of AMIGOS for their support of the research through the *CACD*.

Direct all correspondence to: Anna Perrault, School of Library and Information Science, University of South Florida, 4202 East Fowler Ave. CIS 1040, Tampa, Florida 33620-7800. <perrault@luna.cas.usf.edu>.

Library & Information Science Research, Volume 21, Number 1, pages 47-67

Copyright © 1999 Ablex Publishing Corporation

All rights of reproduction in any form reserved. ISSN: 0740-8188

Research takes the guesswork or estimation out of management or, stated another way, it substitutes data or facts for instinct. When the "serials crisis" arose in the latter half of the 1980s, it was widely believed that academic library collections were suffering serious repercussions as a result (the instinct). In the absence of any detailed standardized data, special studies had to be conducted to obtain a better understanding of how the collections of university libraries were being affected, both by the rise in the price of materials and changes in the environment of academic libraries. The problem of how to study such resources is the subject of this article.

PROBLEM STATEMENT

The primary focus of collection evaluation and assessment has been the determination of strengths and weaknesses in individual library collections as measured against a peer group. The responsibility for the collecting and analysis of local or regional data lies with the immediate parties involved.

Network management reports are routinely compiled for certain categories of library operations, for example, circulation, bibliographic control of library holdings, interlibrary loan transactions, and number of searches by terminal location. Data sufficiently detailed for collection analysis and assessment are usually not supplied in routine management reports. If data for national comparisons are what is desired, readily available data-sets and statistical series are not sufficiently detailed to conduct research on national acquisitions and collecting patterns. There have been only a few reported studies of collecting patterns on a national scale. These studies all used data drawn from the OCLC Online Union Catalog or the *OCLC/AMIGOS Collection Analysis* system.

The methodological problem is represented in the need to study the collective resources base of academic libraries on a national scale. This paper discusses the challenges and limitations of content analysis of national library peer groups utilizing the *OCLC/AMIGOS Collection Analysis CD*. Instead of the comparison of one individual research collection to a peer group of libraries, the collective or aggregated resources base of national peer groups of libraries are studied in this article.

SIGNIFICANCE OF THE PROBLEM

Research on national collecting patterns can produce profiles of national peer groups of libraries and data on collecting patterns over time. These profiles and collection patterns can be utilized for comparison in performing regional or local library collection assessments. The findings of these studies are required in conjunction with the more general data of standard statistical series to track national trends in acquisitions expenditures and the effects of budgetary and economic environmental factors on the collective resources base of academic

libraries. Measurement of such collective resources is integral to the improvement of library performance.

The collection parameters of national peer groups will serve well as proxies indicating the collective resources which are available for resource sharing. The access philosophy of library service, the current paradigm, makes it imperative that the collective resources base be understood in order that acquisitions decisions be made within the context of both local needs and the larger context of network access.

REVIEW OF RELATED RESEARCH

The emphasis in this review of related research is on reported studies of national collecting trends. Broader collection studies, descriptive articles, and specific collection evaluation projects are included in a comprehensive bibliography by Nisonger (1992).

The majority of the collection analysis studies on groups of library collections concentrate on the findings of overlap and duplication of titles among collections. These studies are, by and large, for monographic holdings, although a few studies include serials.

Potter (1982) reviewed the early research in overlap and diversity of library collections and placed the studies into four categories:

- Studies related to union catalogs;
- Studies related to centralized processing;
- Studies for collection development; and
- Studies undertaken to further understanding of collections.

The first three of these categories all have the immediate application of results to practice as their rationale. The last category is further removed from practice and closer to basic research—the establishment of a knowledge base for the understanding of collecting patterns in general, rather than the examination of particular local or regional groups of collections.

The purposes for the study of groups of library collections have altered since Potter established the above four categories. The focus of studies for union catalogs and centralized processing centers was the creation of bibliographic records and centralized bibliographic control. The emphasis was on the sharing of bibliographic information. The resource sharing/access paradigm has been the impetus behind the majority of studies of groups of library collections since the mid-1980s. The emphasis in these studies is on the diversity of collective resources for meeting the needs of users within a group of library collections. The rationale is to discover who owns what in order that the collective resources can be shared.

Medina (1995) conducted a thorough review of the literature of overlap studies with coverage through 1993. Her review includes tables synthesizing and catego-

rizing the major findings of the overlap studies. Widely varying methodologies for types of libraries, size of sample, and other variables make any summary of the findings of overlap studies difficult. The findings of overlap studies on local or regional groups of libraries form only a partial basis for comparison with the findings of national scale studies.

There have been only a small number of studies which have researched national collecting patterns and the composition of academic library collections in the aggregate. With the exception of the first pioneering, now classic study of the Association of Research Libraries (ARL) by Merritt (1942), these studies have all been conducted in the 1990s at a point when the national utility databases had reached a size and breadth reflective of the majority of academic libraries' holdings.

University Libraries and Scholarly Communication (Cummings, 1992) reports a study which has had wide influence on academic library research. The study, funded by the Mellon Foundation, was conducted to analyze trends in the volume of acquisitions, the worldwide pool of publications from which acquisitions are made, prices of journals and monographs, and library expenditures. The "principal objective of this research" was to "describe the library landscape as it appears today, in its collecting, operating, financial, and electronic dimensions" (Cummings, p. i) The study used the Association of Research Libraries' database of library statistics as its main data source and focused on 24 ARL libraries for close examination. Other data utilized were book production statistics and overall university expenditures from the National Center for Education Statistics. The findings of the study all indicated that a retrenchment in collections was taking place in academic libraries.

In the "Foreword" of the Mellon study on university libraries, the hope is expressed that others will be interested "in pursuing in more detail questions that we have only highlighted" (Cummings, p. i). Subsequently, the American Association of Universities (AAU) and the ARL, with funding from the Mellon Foundation, sponsored the AAU/ARL Foreign Acquisitions Project which was directed toward developing a clearer understanding of the forces influencing North American research libraries' ability to build and maintain collections of publications produced outside the United States and Canada. The Project, coordinated by Jutta Reed-Scott, gathered data through the work of 14 country specific task forces composed of area studies librarians; through surveys of bibliographers, surveys of vendors supplying overseas materials; analyses of national cataloging data (from OCLC) for the period 1988-1994; and five sampling studies. A number of the task forces were under the auspices of area studies associations or committees.

The results of the four-year AAU/ARL Foreign Acquisitions Project study were analyzed in *Scholarship, Research Libraries, and Global Publishing*, authored by Reed-Scott (1996). The detailed studies undertaken in the project "provide persuasive evidence of a growing gap between the level of acquisitions of overseas materials and the explosion of global knowledge" (Reed-Scott, p. 5). A number of

recommendations were made centered around the organization of a “distributed program for ‘access’ to global resources” (Reed-Scott, p. 145).

The AAU/ARL study concentrated on foreign materials in U.S. and Canadian research libraries through an area studies approach to publications by country of origin. Although it utilized quantitative data supplied by OCLC, the study was primarily qualitative. Research by Perrault (1994a, 1994b, 1995) was being conducted concurrently with the AAU/ARL Foreign Acquisitions Project. It is the first reported research to study the ARL library collections as an aggregated resources base by both subject and language grouping parameters. The data were extracted from the 1991 *OCLC/AMIGOS Collection Analysis CD* by subject and language of publication rather than origin of publication.

The research by Perrault was designed to determine the effects of the decline in book acquisitions on the collective resources base of ARL libraries for the time period 1985-1989. Her findings were complementary to those of the AAU/ARL study but more extensive and detailed.

The study by Perrault established that there was an increase in homogeneity in the acquisitions of the group of member libraries of the ARL in 1989 from that of 1985. There was a decrease in foreign language acquisitions, a decrease in the percentage of unique titles to the total in many subject areas, and an increased concentration on core materials, indicating less diversity and more homogeneity in academic library collections in the future.

A series of studies on the acquisitions decline in ARL libraries by interdisciplinary fields was conducted by Schwartz (1992a, 1992b, 1993, 1994). He devised a model to measure literature loss which he defined as the “growing gap between book publication output and the aggregate holdings of libraries in the Association of Research Libraries” (1993, p. 1151). The methodology was to extract data from both the OCLC Online Union Catalog database and the *OCLC/AMIGOS CACD*. Schwartz utilized subject searching of the Union Catalog to establish a universe of publication. The holdings of the ARL libraries as reflected in the *CACD* were compared to the universe of publications found in the Online Union Catalog. In the subject areas of anthropology, Judaic studies, international relations, and psychology, Schwartz found literature loss on the same order of magnitude in all four fields. While annual book output has continued to rise, the average collection in the largest ARL libraries holds only half of the mainstream literature with smaller research libraries holding only one-fourth (1994, p. 136).

Perrault and Schwartz are the only researchers who have studied national collecting patterns by imprint year. The studies by Perrault, Reed-Scott, and Schwartz were all performed on the collective resources base of ARL libraries and all concerned the retrenchment in academic library collections commencing in the 1980's.

The collective resources base of 427 college libraries in the size range of 100,000-299,999 titles was the subject of a study by Hardesty and Mak (1994). The study addressed the research question “Does a core collection exist among undergraduate libraries?” The *OCLC/AMIGOS CACD* was the primary source of

data. There was a high degree of diversity among the collections and the core collection among the smaller academic libraries “appears to be quite small—only 76 titles in the 1.8 million-record *CACD* database were held by 75 percent or more of the smaller academic libraries investigated” (p. 366).

This article continues research on national collecting patterns and adds to the findings of the studies reviewed here.

METHOD

Collection analysis is a type of bibliometric research. It is content analysis of library collections studying the variables of collection age, monograph/serial ratios, English/non-English language ratios, collecting patterns of foreign language materials, proportions of collections or collecting patterns by subject areas, overlap and diversity in groups of library collections.

Procedures

There are several options for consortia and networks to obtain collection analysis data and reports. One option is that of designing custom collection analysis programs to extract the data from local or regional network databases. Another option is to use the services of a bibliographic utility such as OCLC or WLN for customized reports. Both utilities offer services which produce collection analysis reports for individual libraries and consortia. Collection analysis research for national collecting patterns can be conducted through the extraction of data from the bibliographic utility databases and customized products.

Data from standard statistical series produced by the ARL, the Association of College and Research Libraries, the Association of Library Collections and Technical Services *North American Title Count* (1989,1993,1995,1997), and the United States Department of Education’s National Center for Education Statistics can also be utilized in analyzing and interpreting the national collection analysis findings. Statistics on publication trends from published sources and vendor reports are other sources of supplemental data utilized in the study of national collecting patterns.

The following analyses of national collective resources are based on data derived from the *OCLC/AMIGOS Collection Analysis CD*¹ product, which contains a subset of the OCLC Online Union Catalog database.

¹ The *OCLC/AMIGOS CACD* has been the subject of a number of descriptive articles and utilized in reports of collection analysis research. The advantages and disadvantages of using the *CACD* have been set forth in a number of articles (Dillon, 1989; Joy, 1992; Radke, 1991; Wood, 1996). Sanko (1997) has compiled a bibliography of studies utilizing the *CACD*.

While the purpose of the *OCLC/Amigos Collection Analysis* system is the comparison of one library collection to a peer group, the *CACD* can be utilized to study the aggregated collections of the standard library peer groups. Instead of the comparison of one individual collection to a peer group of libraries, the collective or aggregated resources base of national peer groups of libraries are studied. Using the data in this collective manner enables the researcher to examine collecting patterns for an entire group of libraries and to formulate conclusions about the collective resources base of that group of libraries. The data can be extracted and analyzed by imprint year, subject, and language categories; all these variables can be cross tabulated.

Methodological Assumptions

- The bibliographic records for non-serial imprints in the *OCLC/AMIGOS Collection Analysis CD* system represent the acquisitions of the libraries whose holdings are in the database for a ten year time period.
- The data from the *CACD* on number of imprints by year correspond closely to acquisitions numbers for that same year.

Limitations

- The database is limited to titles published within an a ten year span. Titles acquired and/or cataloged but not published within the ten year span are not included in the database.
- The data include monographic imprints for the ten years *cataloged* with holdings in OCLC by the libraries and thus represent cataloged imprints no matter how acquired. That is, the materials were not necessarily purchased but could have been acquired through gift, exchange, etc.
- All publications not having a Library of Congress or National Library of Medicine call number in the bibliographic record are excluded from the database. Theses and dissertations are excluded. U.S. government and state government publications are excluded. Thus, every imprint fitting the time span and cataloged by the libraries in the peer groups may not be included in the *CACD* database (*OCLC/AMIGOS, User Guide 1991*).
- Unique titles are determined by the holding symbol for only one library in a peer group being attached to a bibliographic record. A title which is unique in one peer group can also be owned by libraries in another peer group. Thus from a resource sharing point of view, the title may not be unique among all libraries in the OCLC Union Catalog database but only "uniquely held" within one peer group.

Confounding Factors. In addition to the parameters set for inclusion of bibliographic records in the *CACD*, changes over the years have resulted in confounding factors in using the product for research purposes. The number of

libraries for which bibliographic records are included in the *OCLC/AMIGOS CACD* database has changed from one edition to another. Prior to 1995 the product only included the holdings of academic libraries. In 1995, the database was expanded by more than 200,000 titles with the inclusion of law and health science libraries; junior, community and vocational colleges; public libraries, as well as additional academic libraries. In 1995, the database contained approximately 2.1 million abbreviated bibliographic records and aggregate holdings information extracted from the OCLC Online Union Catalog. Prior to 1993 only seven languages were included, but the language possibilities were expanded to select any language included in the OCLC database.

There was also a change in the number and timeliness of imprint years included. Because the purchasers of the *CACD* were utilizing it for current selection, they wanted the product to be as timely as possible. Beginning in 1995, the elapsed time before an imprint year could be included in the ten year database was moved forward by one imprint year. Instead of each edition ending with a two year gap in coverage from the last imprint year included to the year the product was constructed, the gap was reduced to six months with the previous calendar year being the last included in the database.

All of these changes in the parameters of inclusion for the product mean that data obtained from the *CACD* for a specific measure will not be equivalent numbers for the same year from one edition of the *CACD* to another.

To illustrate the point about the difference in numbers obtained from different editions of the *CACD*, Table 1 shows the total number of titles for a five-year period from three different editions of the *CACD*.

The percentages under the totals for 1996 and 1997 show the increase in the number of records for the same imprint year in subsequent editions of the *CACD*. It is obvious that the more current the imprint year, the smaller the number of records in the database. It can be assumed that cataloging lag is a factor in the lower number of records for more current imprint years. For the 1997 edition, the increase in records is at a lower rate than for the 1996 edition. But the number of libraries included is not the same in each year.

TABLE 1
Difference in Number of Titles by Imprint Year in 1995-1997
Editions of CACD for Peer Group 1

| | 1990 <i>imprints</i> | 1991 <i>imprints</i> | 1992 <i>imprints</i> | 1993 <i>imprints</i> | 1994 <i>imprints</i> |
|--------------------------|-----------------------------------|-----------------------------------|------------------------------------|------------------------------------|------------------------------------|
| 1995 CACD (# lib.=80) | 170,495 | 168,039 | 160,409 | 138,490 | 89,003 |
| 1996 CACD (# lib.=87) | 182,819 (7% more than 1995) | 183,224 (9% more than 1995) | 184,161 (15% more than 1995) | 177,155 (28% more than 1995) | 160,048 (80% more than 1995) |
| 1997 CACD (# lib.=95) | 191,960 (5% more than 1996) | 194,472 (6% more than 1996) | 197,731 (7% more than 1996) | 196,930 (11% more than 1996) | 187,947 (28% more than 1996) |

The differences in the data from one edition of the *CACD* to another illustrate the necessity of consistently using data from only one edition within the same series analysis. The data should not be mixed with one time series analysis utilizing data drawn from different editions of the *CACD*. The point in the research on national collecting patterns is to discover trends over time, but the data do need to be gathered systematically and consistently.

Data Collection and Analysis. For this research, the data extraction from the *CACD* was performed via Microsoft Access. Queries were run to extract data according to the parameters of imprint year by Library of Congress class categories and by language groupings. The data obtained from the queries are then imported into Microsoft Excel to construct spreadsheets of the results of the queries. Summary tables for the analysis are constructed from the results of the spreadsheet calculations.

EXAMPLES OF FINDINGS ON NATIONAL COLLECTING PATTERNS

National collecting patterns are the composite results of the acquisition of materials by individual libraries. There are very little data reported in standard statistical series on the acquisitions rates for libraries. The first example of findings on national collecting patterns looks at the feasibility of using data from bibliographic utilities, that is cataloging records, as a reflection of acquisitions patterns.

TABLE 2
Monograph Acquisitions in ARL Libraries 1986-1995

| <i>Year</i> | <i>Col. 1 Median Price \$</i> | <i>Col. 2 Median Price Ann. % +/-</i> | <i>Col. 3 Expend. Ann. % +/-</i> | <i>Col. 4 Avg. # Titles Purchased</i> | <i>Col. 5 Titles Purch. Ann. % +/-</i> |
|-------------|---|---|--|---|--|
| 1986 | 28.65 | | | 32,210 | |
| 1987 | 31.76 | 11 | -5 | 27,214 | -18 |
| 1988 | 35.63 | 12 | 7 | 26,541 | -2 |
| 1989 | 37.74 | 6 | 9 | 27,268 | 3 |
| 1990 | 40.26 | 7 | 7 | 27,999 | 3 |
| 1991 | 42.04 | 4 | 5 | 28,027 | 0.1 |
| 1992 | 43.31 | 3 | 3 | 27,158 | -3 |
| 1993 | 41.78 | -4 | -4 | 25,583 | -6 |
| 1994 | 44.51 | 7 | 1 | 25,803 | 0.86 |
| 1995 | 45.27 | 2 | 4 | 25,719 | -0.33 |

Source: ARL Statistics, 1994/95
OCLC/AMIGOS CACD, 1996 ed.

Acquisitions in ARL Libraries

The median price of monographs as acquired by the ARL libraries and the annual increase or decrease in the median price are shown in the first two columns of Table 2. Column 3 shows the annual increase/decrease in the average expenditures by the ARL libraries for monographs in the same years. The last two columns show the average number of monographs purchased for those same years by the 63 libraries which reported that data element. The average number of monographs purchased for a single year reports all purchases regardless of the imprint year of those purchases. The last column shows the annual increase/decrease in the average number of monographs purchased. These data are reported on an annual basis by the ARL libraries.

It is readily apparent that with the exception of 1991, the median price of monographs rose every year. With the exception of 1987, 1992, and 1993, expenditures for monographic acquisitions also rose each year. Although the expenditures rose, the number of monographs purchased continued to fluctuate within a downward trend throughout the 10-year period. The difference between the average number of monographs purchased in 1986 and in 1995 is a decrease of 22 percentage points.

To add to this analysis of the trend in monographic acquisitions, Table 3 shows the average number of titles per ARL library in the *CACD* database compared with the number of monographs purchased each year.

The same data from Table 2 reported for average number of monographs purchased by ARL libraries appears in the first column of Table 3. The second column shows the average number of titles per library per imprint year for the ARL group as reflected in the *CACD* database. The third column shows the aver-

TABLE 3
Monograph Acquisitions in ARL Libraries Compared with
Number of Titles in CACD

| | <i>Col. 1</i> | <i>Col. 2</i> | <i>Col. 3</i> |
|-------------|--------------------------------|--|---|
| <i>Year</i> | <i>Avg. # Titles Purchased</i> | <i>Avg. # Titles per imprint yr. in CACD</i> | <i>Titles in CACD as % Titles Purchased</i> |
| 1986 | 33,210 | 23,157 | 69 |
| 1987 | 27,214 | 22,842 | 84 |
| 1988 | 26,541 | 23,471 | 88 |
| 1989 | 27,268 | 23,995 | 88 |
| 1990 | 27,999 | 23,967 | 86 |
| 1991 | 28,027 | 23,158 | 85 |
| 1992 | 27,158 | 22,951 | 85 |
| 1993 | 25,583 | 22,870 | 89 |
| 1994 | 25,803 | 21,560 | 84 |
| 1995 | 25,719 | 16,948 | 66 |

Source: ARL Statistics, 1994/95
 OCLC/AMIGOS CACD, 1996 ed.

age number of titles in the second column as a percentage of the number of monographs purchased in the first column.

The average number of titles per library in the *CACD* is consistently in the 80% range as a proportion of the number of monographs purchased each year, with the exception of the first and last years. It possible that with the better funding of 1986, acquisitions were spanning a greater variety of imprint years than acquisitions in later, leaner budget years accounting for fewer titles purchased in that one year. The number of imprints in the last year, 1995, is not a reliable datum as there has not been sufficient time lag between the production of the 1996 edition and the last imprint year included in that edition as established in Table 1.

For the imprint years 1987-1994, there appears to be a relationship between the reported number of monographs purchased and the average number of titles per ARL library per imprint year. For those years, the number of titles per year varies within a narrow 84%-89% range with the reported number of monographs purchased.

There is one more reported data element which can be compared with the data in Tables 2 and 3. Table 4 displays the total number of titles treated annually for the Blackwell approval plan from 1991-1995. (Blackwell North America, *Approval program coverage and cost study, 1991-1995*).

The approval plan treats titles judged suitable for university and college library collections. Table 4 shows the average number of titles per imprint year for ARL libraries from the *CACD* as a proportion of total titles treated for the Blackwell approval plan. Again there are apparent patterns. The number of monographs purchased varies within a 77%-85% range of the number of titles treated on approval. The average number of titles per ARL library for the same years in the *CACD* database varies closely within a lower range of 66.5%-70.5%.

It is *not* being suggested that approval plan purchases make up 70-80% of the ARL library purchases. The interpretation is that the quantity, the number of monographs purchased in a given year, closely corresponds to the number of imprints for that year which have been judged to be appropriate acquisitions for

TABLE 4
Acquisitions as a Proportion of Approval Plan Coverage
ARL Libraries, 1991-1995

| Year | App. Plan | ARL Mono. Purchased** | % Col. 1 | CACD | % Col. 1 |
|------|------------------|--------------------------|----------|---------------------|----------|
| | Total Titles* | | | Avg. # Titles*** | |
| 1991 | 32,843 | 28,027 | 85 | 23,158 | 71 |
| 1992 | 33,186 | 27,158 | 82 | 22,951 | 69 |
| 1993 | 33,220 | 25,583 | 77 | 22,870 | 69 |
| 1994 | 32,414 | 25,803 | 79 | 21,560 | 67 |
| 1995 | 30,662 | 25,719 | 84 | 16,948 | 55 |

Note: *Blackwell's Approval Plan Cost and Coverage

**Data for monographs purchased from ARL Statistics, 1994-95. N=63

***Data for Avg. No. of titles from 1996 edition of the OCLC/AMIGOS CACD. N=87

academic libraries. The interpretation also is that the average number of titles per imprint year in the *CACD* database follows a similar pattern to the approval plan and library acquisitions data. From these data we can anticipate that collecting patterns as found in the analysis of the distribution of bibliographic records in the *CACD* do follow similar patterns as compared with reported acquisitions data.

A few points about the nature of acquisitions can add to the considerations to be borne in mind in interpreting these data. While the data in the tables are all single imprint year numbers, in actuality one year's acquisitions budget buys titles which span a variety of different years of publication. Both methodological assumptions for the research are predicated upon the bibliographic records in the *CACD* corresponding to the acquisitions patterns for the same imprint years. That there is a correspondence between number of titles purchased and number of titles cataloged in the *CACD* database has been shown in Tables 3 and 4.

These findings do raise a number of questions with regard to the use of the cataloging data to study acquisitions patterns. All of the data in Tables 2-4 are from annual statistical report series, with the exception of the data on average number of titles owned per library by year. The last set of data were not gathered on an annual basis but are *ex post facto* from one point in time, the 1996 edition of the *CACD*.

Another way to characterize the data is to say that the annually reported data are fixed. But numbers of bibliographic records in the cataloging databases change—they don't stand still. Libraries constantly add more resources of various imprint years. Thus, the number of bibliographic records per imprint year in the OCLC Online Catalog or the *CACD* can increase over time. Cataloging lag can also be a factor in comparing acquisitions numbers and cataloged holdings. While these data suffice to discern patterns, they are not extremely precise in that the absolute numbers are subject to change and revision in the future. But the point, again, is that the data are sufficiently accurate to reveal patterns or long term trends.

The analysis of the data in Tables 2-4 demonstrate that changes in the product composition and the acquisitions and cataloging speed of various libraries are all confounding factors in the interpretation of cataloging data as corresponding to or reflecting acquisitions trends. These data can be utilized, but the claims in interpretation must not exceed the capacity of the data.

Collecting Patterns for Three Academic Library Peer Groups

Up to this point, the methodology, assumptions, and limitations in conducting research on national collecting patterns have been explored using data on the ARL peer group. This section reviews findings from comparative analysis between peer groups. Tables 5-10 and Figures 2-5 display data for four peer groups from the 1996 *CACD*. The 1996 *CACD* database contains the bibliographic records of 2,497 libraries which comprise Peer Group 14—the database. Three peer groups of academic libraries are studied: the ARL group, Peer Group 1; libraries with collections from 700,000- plus volumes, Peer Group 4; and libraries with collections of

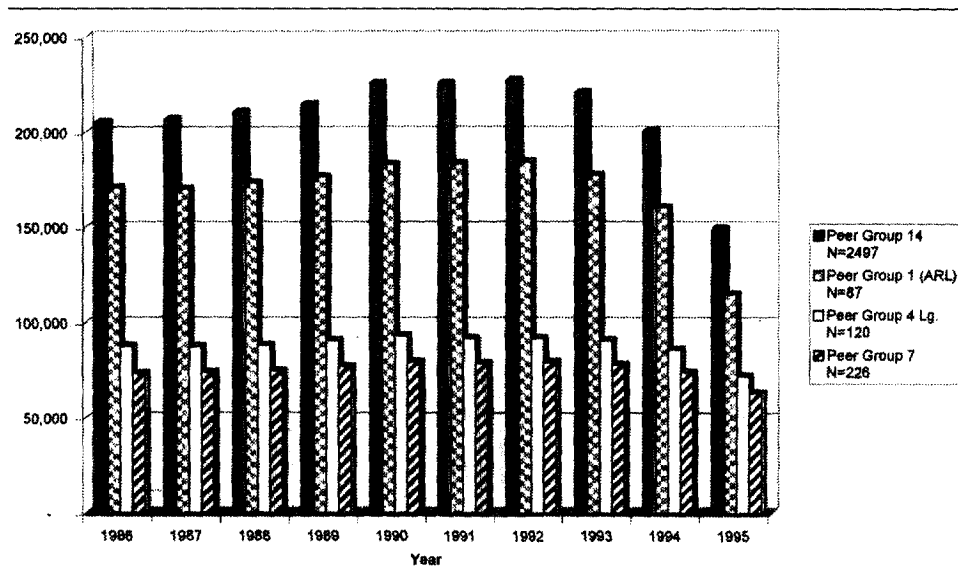
TABLE 5
Number of Bibliographic Records by Peer Group
1996 CACD

| | Peer Group 14 N=2497 | Peer Group 1 (ARL) N=87 | Peer Group 4 Lg. N=120 | Peer Group 7 N=226 |
|------|-------------------------|----------------------------|---------------------------|-----------------------|
| 1986 | 205,218 | 170,719 | 87,589 | 73,287 |
| 1987 | 206,853 | 169,949 | 87,456 | 74,020 |
| 1988 | 210,438 | 173,144 | 88,235 | 74,823 |
| 1989 | 214,455 | 176,304 | 90,732 | 77,076 |
| 1990 | 225,669 | 182,819 | 93,171 | 79,569 |
| 1991 | 225,648 | 183,224 | 91,739 | 78,768 |
| 1992 | 227,202 | 184,161 | 91,865 | 79,401 |
| 1993 | 220,588 | 177,155 | 90,768 | 78,175 |
| 1994 | 200,149 | 160,048 | 86,049 | 74,113 |
| 1995 | 149,221 | 114,675 | 72,130 | 63,126 |

300,000-699,999 volumes, Peer Group 7. All data are derived from the 1996 edition of the *OCLC/AMIGOS CACD*.

The comparative size of collections between the three peer groups of academic libraries and Peer Group 14 can easily be seen in Table 5 and in Figure 1 which graphs the same data. While the ARL peer group has a much smaller number of libraries than peer groups 4 and 7, it makes up the largest block of bibliographic records in the database. Even though the holdings of 2,497 libraries are represented in the database, the bibliographic records of the ARL libraries make up over 80% of the titles for any given imprint year. The number of bibliographic records with

FIGURE 1
Number of Bibliographic Records by Peer Group
1996 CACD



no ARL library location is consistently less than 20% of total records in the *CACD* by imprint year. From these data it can be seen that studying the collecting patterns of ARL libraries can be said to be studying the composition of the "national collection."

The data in Table 5 demonstrate that the patterns of numbers of imprints by year are very similar for all three groups of academic libraries. The smaller libraries have lower numbers, but the pattern is the same, although less pronounced. The total number of different titles for peer groups 4 and 7 are not separated by a wide span in the absolute numbers. When it is borne in mind, however, that the possibility of different titles which could be owned for each imprint year is over 200,000 titles in the database, there could be a wide variety of titles represented by the number for each peer group.

Recalling the estimate made by Haas in 1977 that the combined acquisitions of the ARL libraries involved a total of perhaps 300,000 "discrete" titles, it can be seen from Table 5 that his "guesstimate" was off by roughly 100,000. It has only been in recent years that the data from the OCLC Online Catalog and the *OCLC/AMIGOS CACD* has made it possible to determine such detailed information on library collecting patterns.

The most dramatic shift in collecting patterns in the last decade can be seen in the changes in the ratio of English language to non-English imprints over time.

The proportions of English language titles to non-English language titles are shown in Table 6 for Peer group 1 and Peer group 14, the entire database. The startling thing about these data is that in 1985, the aggregated collections of ARL libraries contained a larger number of foreign language monograph titles than English language titles. From previous research it has been established that by 1989 the balance had swung to predominance of English language materials (Perrault, 1994a, 1994b, 1996).

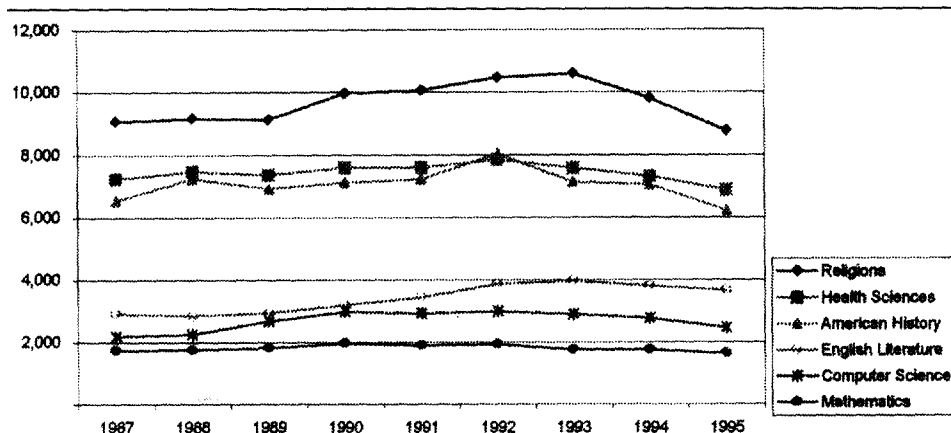
It is understandable that English language materials would predominate in Peer group 14 which is composed of 2,497 libraries, the majority not from research institutions. But the proportion of English to non-English increases considerably for Peer Group 14 between 1985 and 1990.

In 1985, the proportions for ARL libraries were 49%/51%, English to non-English, very nearly equal. In 1990, the ratio was 52%/48% with the balance switching to English language predominance. For peer group 14 in 1985, the ratio was 54%/46%, with an 8 percentage point difference, changing in 1990 to 58%/42%, seeing English gaining in proportion to non-English with a difference of 16

TABLE 6
Proportions by Language

| | <i>Peer Group 1</i> | | <i>Peer Group 14</i> | |
|-------------|---------------------|-------------|----------------------|-------------|
| | <i>1985</i> | <i>1990</i> | <i>1985</i> | <i>1990</i> |
| English | 81,998 | 95,706 | 108,981 | 131,391 |
| Non-English | 85,471 | 87,113 | 92,803 | 94,278 |
| Total | 167,469 | 182,819 | 201,784 | 225,669 |

FIGURE 2
Collecting Patterns by Subject Category
Peer Group 1, ARL Libraries



Source: ARL Statistics, 1994/95
 OCLC/AMIGOS CACD, 1996 ED.

percentage points between the two. This shift in proportions by language is widely attributed to the reduction in the number of monograph acquisitions in research libraries which has been shown in Table 2.

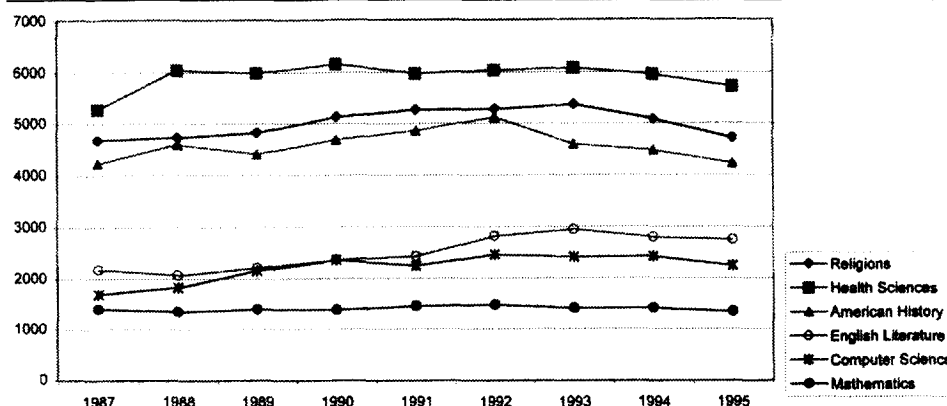
Another comparative analysis tracks collecting patterns by subject categories. In Figures 2-5 and Tables 7-10, six subject categories are displayed for the years 1987-1995 for the four different peer groupings. Figure 2 and Table 7 display the pattern for the ARL Peer Group 1. Figure 3 and Table 8 display data for the same parameters for Peer Group 4, libraries with over 700,000 volumes. Similarly, Figure 4 and Table 9 show the patterns for Peer Group 7, libraries with 300,000-699,999 volumes. Peer Group 14, the entire CACD database, is shown in Figure 5 and Table 10. Although the lines on the graphs appear to be very similar for all subjects and all peer groups, a closer examination reveals both interesting differences and similarities.

TABLE 7
Collecting Patterns by Subject Category
Peer Group 1, ARL Libraries

| | 1987 | 1988 | 1989 | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 |
|--------------------|-------|-------|-------|-------|--------|--------|--------|-------|-------|
| Religious | 9,062 | 9,155 | 9,122 | 9,956 | 10,044 | 10,450 | 10,578 | 9,799 | 8,773 |
| Health Sciences | 7,231 | 7,473 | 7,360 | 7,600 | 7,588 | 7,839 | 7,582 | 7,309 | 6,881 |
| American History | 6,534 | 7,230 | 6,919 | 7,105 | 7,218 | 8,035 | 7,129 | 7,050 | 6,217 |
| English Literature | 2,906 | 2,846 | 2,940 | 3,183 | 3,432 | 3,856 | 3,949 | 3,796 | 3,650 |
| Computer Science | 2,190 | 2,255 | 2,679 | 2,973 | 2,925 | 2,981 | 2,892 | 2,772 | 2,454 |
| Mathematics | 1,729 | 1,748 | 1,815 | 1,946 | 1,889 | 1,917 | 1,751 | 1,747 | 1,618 |

Source: ARL Statistics, 1994/95
 OCLC/AMIGOS CACD, 1996 ED.

FIGURE 3
Collecting Patterns by Subject Category
Peer Group 4, Academic Libraries 700,000+ Volumes



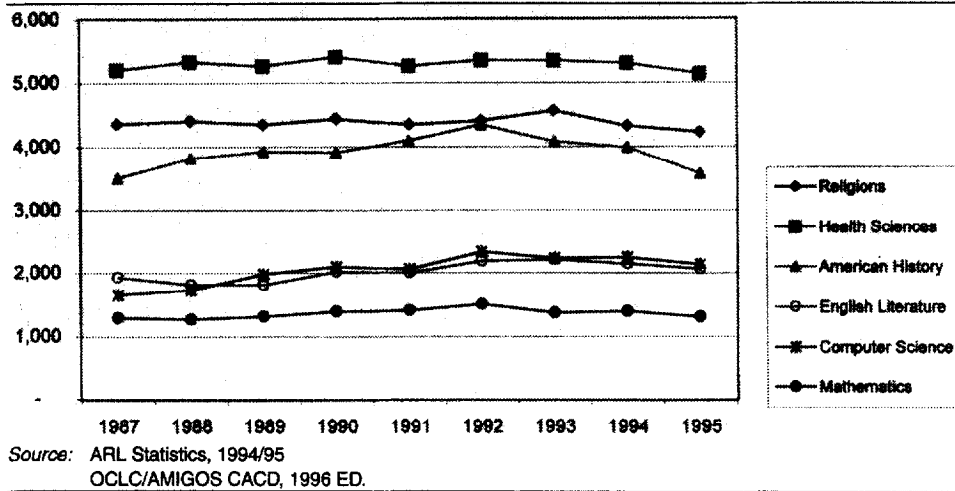
Source: ARL Statistics, 1994/95
 OCLC/AMIGOS CACD, 1996 ED.

The subject category with the largest number of titles in peer groups 14 and 4 is health sciences. Religion vies with health sciences as the top collected category, highest in peer groups 1 and 7. The study of religions is an academic pursuit, whereas the health sciences represent applied education as well as research.² The high number of health sciences titles in all four peer groups shows the prevalence of medical and allied health programs at all sizes of higher education institutions.

The number of titles acquired in the six subject categories corresponds with the available universe of publication for those subjects. For example, math (QA excluding computer science) has a very similar pattern of acquisitions for all three peer groups with the absolute numbers not differing widely when the size of collections is taken into consideration. It would seem, from the similarity in the numbers, that there is indeed a small, tight universe of publication in mathematics, and that the majority of academic libraries are acquiring a large proportion of that universe. This conjecture is supported by approval plan data (Blackwell, 1991-1995). The number of titles treated for the approval plan in the health sciences ranges from 3,500 to 4,000 from 1991-1995. In mathematics the number ranges between 1,600 and 2,500. Just as the number of titles for the subject areas in the CACD per year fluctuate within a narrow range, so do the number of titles treated

² Separate medical school collections of the ARL libraries may not be accounted for under the main library location symbol. Thus the health sciences data for the ARL group may not include separate medical school collections which may appear in peer group 14 as separate libraries. Smaller universities in which the total collections are represented in the CACD and which have allied health programs would have a more accurate reflection of those collections because all holdings appear under the main library symbol.

FIGURE 4
Collecting Patterns by Subject Category
Peer Group 7, Academic Libraries 300,000-699,999 Volumes



on approval. The pattern for both the approval plan and the records in the *CACD* are very similar. The larger number of titles in the *CACD* peer group analysis is an indication that acquisitions are coming from other sources in addition to the approval plans.

In contrast to mathematics, religion is a category with a very large universe of worldwide publication. Titles in religion rank high for collecting in all four peer groups.

FIGURE 5
Collecting Patterns by Subject Category
Peer Group 14, the CACD Database

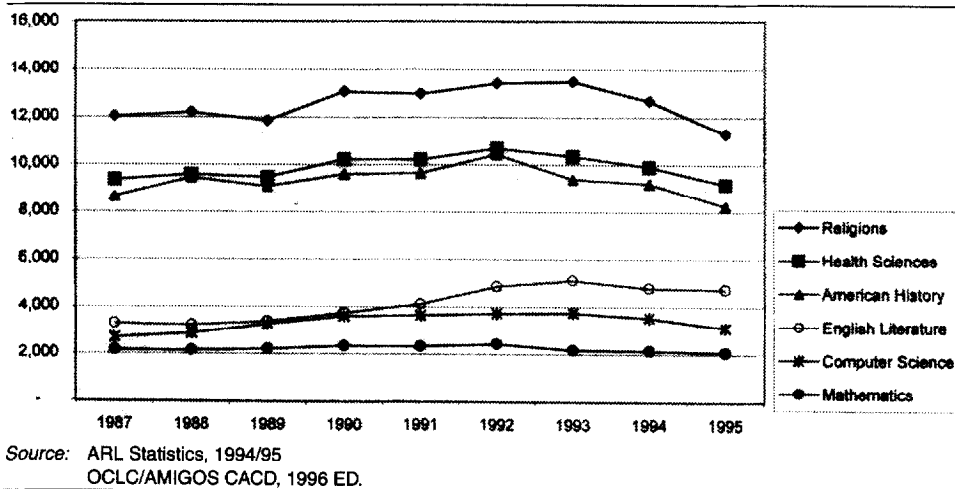


TABLE 8
Collecting Patterns by Subject Category
Peer Group 4, Academic Libraries 700,000+ Volumes

| | 1987 | 1988 | 1989 | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 |
|--------------------|------|------|------|------|------|------|------|------|------|
| Religious | 4679 | 4738 | 4824 | 5130 | 5267 | 5278 | 5366 | 5081 | 4709 |
| Health Sciences | 5280 | 6045 | 5991 | 6162 | 5978 | 6038 | 6078 | 5949 | 5715 |
| American History | 4228 | 4601 | 4409 | 4688 | 4862 | 5121 | 4596 | 4476 | 4222 |
| English Literature | 2173 | 2058 | 2205 | 2359 | 2426 | 2816 | 2941 | 2792 | 2741 |
| Computer Science | 1684 | 1821 | 2149 | 2350 | 2244 | 2461 | 2412 | 2420 | 2238 |
| Mathematics | 1392 | 1340 | 1390 | 1377 | 1451 | 1474 | 1404 | 1404 | 1336 |

Source: ARL Statistics, 1994/95
 OCLC/AMIGOS CACD, 1996 ED.

Other categories with high publications numbers are English literature(PR) and U.S. history (E). The recovery curve for both of these fields displays a similar pattern. Acquisitions appear to be slower or further back of the current imprint year than other subjects, especially the sciences. For this reason, major subjects in the humanities and social sciences appear to have "lost out" in the budget crunch of the 1980's. While there is a natural tendency to attempt to fill in titles missed in bad budget years, the press of current publication does not allow for much retrospective catching up. From the patterns found in these analyses, those subjects which lose out on a current basis, do not appear to be making up for lost time on a retrospective basis. While the trend lines do not evidence volatile swings in any subject category or in the overall numbers, there can be long range impact of small increases or decreases on an annual basis over time.

The data from the CACD alone are not sufficient to form judgments about the underlying rationale for the collecting patterns found in this research. Supplemental data from book publication statistics and approval plan statistics can assist in further interpretation of the data derived from the CACD.

TABLE 9
Collecting Patterns by Subject Category
Peer Group 7, Academic Libraries 300-700,000 Volumes

| | 1987 | 1988 | 1989 | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 |
|--------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Religious | 4,360 | 4,400 | 4,345 | 4,432 | 4,344 | 4,405 | 4,560 | 4,319 | 4,224 |
| Health Sciences | 5,201 | 5,327 | 5,262 | 5,404 | 5,259 | 5,349 | 5,342 | 5,302 | 5,140 |
| American History | 3,507 | 3,817 | 3,922 | 3,909 | 4,096 | 4,344 | 4,081 | 3,986 | 3,574 |
| English Literature | 1,931 | 1,816 | 1,811 | 2,021 | 2,003 | 2,187 | 2,211 | 2,146 | 2,064 |
| Computer Science | 1,654 | 1,729 | 1,981 | 2,101 | 2,059 | 2,336 | 2,233 | 2,247 | 2,135 |
| Mathematics | 1,291 | 1,268 | 1,314 | 1,392 | 1,404 | 1,505 | 1,369 | 1,393 | 1,307 |

Source: ARL Statistics, 1994/95
 OCLC/AMIGOS CACD, 1996 ED.

TABLE 10
Collecting Patterns by Subject Category
Peer Group 14, the CACD Database

| | 1987 | 1988 | 1989 | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 |
|--------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| Religious | 12,021 | 12,193 | 11,857 | 13,073 | 12,964 | 13,410 | 12,489 | 12,672 | 11,307 |
| Health Sciences | 9,359 | 9,580 | 9,474 | 10,221 | 10,206 | 10,685 | 10,353 | 9,907 | 9,171 |
| American History | 8,653 | 9,447 | 9,076 | 9,604 | 9,338 | 10,433 | 9,388 | 9,194 | 8,237 |
| English Literature | 3,287 | 3,213 | 3,369 | 3,727 | 4,106 | 4,846 | 5,085 | 4,765 | 4,718 |
| Computer Science | 2,683 | 2,837 | 3,251 | 3,585 | 3,645 | 3,722 | 3,745 | 3,526 | 3,111 |
| Mathematics | 2,154 | 2,136 | 2,197 | 2,327 | 2,326 | 2,421 | 2,168 | 2,143 | 2,073 |

Source: ARL Statistics, 1994/95
 OCLC/AMIGOS CACD, 1996 ED.

CONCLUSION

These findings demonstrate that it is feasible to study the composition of the collective resources base of academic libraries using data extracted from bibliographic utilities. After a period of three to four years, data appear to be reliable for studying acquisitions and collecting patterns over time. Trends in collecting patterns in the years 1985-1996 which can be discerned from the findings presented in this article are that the:

- Aggregated resources of academic libraries in the United States are becoming increasingly composed of English language publications;
- Budget crisis of the 1980s is still visible in reduced numbers of titles for those years across all subject areas;
- Smaller academic libraries have more stable collecting patterns indicating that the larger research institutions were harder hit by the serials crisis in the 1980s; and
- Number of titles for all categories rises in the 1990s indicating somewhat of a recovery from the decline in the 1980s.

Research on the composition of the national collection should continue. The main source of data for the research on national collecting patterns conducted in the 1990s has been the *OCLC/AMIGOS CACD* which has been discontinued. It is recommended that networks and consortia devote more attention to the provision of collection analysis data as a component of standardized management reports.

The findings of this research demonstrate that collection analysis on a national scale can reveal patterns of collecting by subject and language which contribute to the base of knowledge for resources sharing and local collecting decisions. These collecting patterns are germane to the access paradigm which is currently the prevalent mode in the provision of resources to library users.

REFERENCES

- American Library Association, Association for Library Collections & Technical Services (1989, 1993, 1995, 1997). *National title count*. Chicago, IL: The Association.
- Association of Research Libraries (1985-1996). *ARL statistics*. Washington, DC: The Association.
- Blackwell North America, Inc. (1991-1995). *Approval program coverage and costs study*. Lake Oswego, OR: Blackwell Group.
- Cummings, Anthony M., Witte, Marcia L., Bowen, William G., Lazarus, Laura O., & Ekman, Richard H. (1992). *University libraries and scholarly communication: A study prepared for the Andrew M. Mellon Foundation*. Washington, DC: Published by the Association of Research Libraries for the Andrew W. Mellon Foundation.
- Dillon, Martin (1989). Collection analysis CD: A new approach to collection assessment. *Library Hi Tech News*, 65, 3-5.
- Haas, Warren J. (1977). Organizational structures to meet future bibliographic requirements. *Library Quarterly*, 47(3), 274-84.
- Hardesty, Larry, & Mak, Collette. (1994). Searching for the Holy Grail: A core collection for undergraduate libraries. *Journal of Academic Librarianship*, 19(6), 362-371.
- Joy, Albert H. (1992). The *OCLC/AMIGOS Collection Analysis CD*: A unique tool for collection evaluation and development. *Resource Sharing and Information Networks*, 8(1), 23-45.
- Medina, Sue O. (1995). Duplication and overlap among library collections: A chronological review of the literature. In Thomas W. Leonhardt (Ed.), *Advances in collection development and resource management* (v.1, pp.1-60). Greenwich, CT: JAI Press.
- Merritt, Leroy C. (1942). The administrative, fiscal, and quantitative aspects of the regional union catalog. In Robert B. Downs (Ed.), *Union catalogs in the United States*, Chicago, IL: American Library Association.
- Nisonger, Thomas E. (1992). *Collection evaluation in academic libraries: A literature guide and annotated bibliography*. Englewood, CO: Libraries Unlimited.
- Perrault, Anna H. (1994a). The changing print resource base of academic research libraries in the United States: A comparison of collecting patterns in seventy-two ARL academic libraries of non-serial imprints for the years 1985 and 1989. Unpublished doctoral dissertation, Florida State University.
- Perrault, Anna H. (1994b). The shrinking national collection: A study of the effect of the diversion of funds from monographs to serials on the monograph collections of research libraries. *Library Acquisitions: Practice & Theory*, 18(1), 3-22.
- Perrault, Anna H. (1995). The changing print resource base of academic research libraries in the United States. *Journal of Education for Library & Information Science*, 36(4), 1-16.

- Potter, William Gray. (1982). Studies of collection overlap: A literature review. *Library Research*, 4, 3-21.
- OCLC, Inc. (1991). *OCLC/AMIGOS Collection Analysis CD User Guide*. Dublin, OH: OCLC Online Computer Library Center.
- Radke, Barbara, A. (1991). OCLC/AMIGOS Collection Analysis CD. *CLC Micro*, 7(5), 28-30.
- Reed-Scott, Jutta (1996). *Scholarship, research libraries, and global publishing*. Washington, DC: Association of Research Libraries.
- Sanko, Shannon. (1997). *Electronic collection evaluation selective bibliography*. Dallas, TX: AMIGOS.
- Schwartz, Charles A. (1992). Literature loss in anthropology. *Current Anthropology*, 33, 315-317.
- Schwartz, Charles A. (1992). Literature loss in international relations. *PS: Political Science & Politics*, 25, 720-723.
- Schwartz, Charles A. (1993). Gap analysis of book publication output and aggregate research collections in psychology. *American Psychologist*, 48, 1151-52.
- Schwartz, Charles A. (1994). Empirical analysis of literature loss. *Library Resources & Technical Services*, 38(2), 133-138.
- Wood, Richard J. (1996). Building a better library collection: The CACD way. *Library Software Review*, 15, 22-23.