Using local citation data to relate the use of journal articles by academic researchers to the coverage of full-text document access systems

Jacobs, N., Woodfield, J. and Morris, A.

n.a.jacobs@lboro.ac.uk

j.woodfield@lboro.ac.uk

a.morris@lboro.ac.uk

Department of Information Science

Loughborough University

LOUGHBOROUGH

Leicestershire

LE11 3TU

Abstract

The methodology and findings are presented of an empirical study comparing local

citation patterns with the holdings lists of a number of sources of journal articles. These

sources were the British Library Document Supply Centre (BLDSC) and the BL 'inside'

service, library holdings, ProQuest Direct, SearchBank, EiText and a linking system

including both the Geobase database and the BLDSC. The value of local citation figures

is discussed, as is the concept of a 'core' of journal titles, from both theoretical and

practical perspectives. Using these figures to represent the local use of journal articles,

the coverage of the document sources was found to vary widely. Unsurprisingly, the

BLDSC was found to offer the widest coverage. Newer, electronic systems generally

fared less well, but may offer other advantages.

Introduction

One role of the academic library is creating and managing the connections between local researchers and sources of the scholarly literature[a]. To do this, academic library managers need information both about the scholarly literature that researchers use and about the relevant attributes of possible sources of that literature. If consideration is limited to academic journals, an important medium in most disciplines, then one factor that relates both to what researchers use and to sources of scholarly literature is that of journal coverage. Because there is a practical limit to the number of literature sources that libraries can efficiently and effectively support, whether in the form of subscriptions or of subsidizing document delivery for users, then those sources that most closely match what local researchers use in terms of coverage will be favoured.

The objectives of this paper are, firstly, to evaluate one particular method of assessing journal coverage (that of comparing such coverage with citations) and, secondly, to use this method to assess the potential coverage for academic researchers of a number of sources of journal articles. This statement of objectives implies that there are four major issues to be addressed:

1. The identification of data to indicate local researchers' use of journals.
2. The identification of data to indicate journal coverage.

---

a. The definition of "sources of the scholarly literature" is discussed in the section concerned with the empirical part of this study.

3. A comparison of (1) with (2).

4. An evaluation of procedures (1) – (3).

These four issues are pursued in the four main sections of this paper. The empirical work described in this paper formed a part of the UK eLib project FIDDO; an acronym for Focused Investigation of Document Delivery Options [1]. The FIDDO project was funded by the Joint Information Systems Committee of the UK Higher Education Funding Councils.

Local use of journals

User requirements, use and citations

In 1974, Line [2] provided a succinct summary of the differences and similarities between need, want, demand, use and requirement. Although controversial (his definition of 'demand' does not mention the ability to pay, which renders it unhelpfully distinct from economic definitions of the term), Line's schema reminds us to pay attention to what we are attempting to measure. This paper is concerned with the relationship between the journal coverage of defined sources and what specific defined groups of academic researchers use. It is certain that such use is related in some way to user requirements, but such requirements are not directly measurable. Indeed, it may be inevitable that attempts at such measurement are partially constitutive of that which they purport to measure [3]. The existence of user requirements is inferred from evidence such as reference interviews between librarians and researchers, which are often

4

characterized as interactions that aim to reveal implicit requirements rather than merely answer explicit queries. Whether such implicit requirements can properly be said to exist prior to the reference interview and, if so, whether this model of user requirements also holds for journal coverage (the focus of this paper) are, at best, uncertain. However, those charged with the responsibility of enabling academic researchers to have adequate access to the journal literature, without exceeding a budget, need some means of assessing what researchers would use, were they fully informed of both their own needs and the journals available. It is this concept of 'potential use', rather than the ineffable 'requirements', that is the focus of this paper. Various means have been suggested for its estimation, including asking the researchers directly [4], surveying preferences [5], and citation analysis [6, 7, 8]. Each of these methods aims to generate a list, or lists, of journal titles that can represent what researchers would use, given certain assumptions. The collection management use to which these lists are put is often described in terms of journal cancellation exercises but, given the terms of the access / holdings debate [9, 10], might be better described in terms of selecting the best sourcing option for each title. This paper is concerned with evaluating one of the techniques mentioned above, citation analysis, in terms of its practical value to librarians in making such selections.

Of the several methods of generating journal title lists to represent what researchers would use, given certain assumptions, citation analysis would appear to need justifying, both in terms of those assumptions and in terms of the apparent relative complexity of the approach.

5

Library managers could infer that citation data represented the journals that researchers would use, were they (researchers) fully informed of both their own needs and the journals available, if the following five conditions were met:

(i)   that the citation data related to citations made by the researchers in question;

(ii)   that the researchers were fully informed of their own needs and the journals available to them when they made the citations;

(iii)   that the researchers acted rationally on this knowledge and used those journals that they needed;

(iv)   that citations to journals were isomorphic with use of those journals;

(v)   that past use were a perfect predictor of future use.

In the real world, as implied by Line [11], none of these conditions can be met completely.

With respect to condition (i), some analysts [12] have been content to work with global or national citation figures such as the Institute for Scientific Information's (ISI) Journal Citation Reports, whereas others [13, 14] have taken a more local focus, looking at the citations of researchers working in a specific institution or discipline. The latter approach is taken in this paper. Kneider [15] compared the two approaches, confirming Line's assertion [16] that such correlations get weaker as the citation frequency diminishes. The use of local, rather than global, citation counts certainly defends the technique against the charge that it is based on figures that have little relevance to local circumstances.

Condition (ii), regarding the knowledge states of researchers when they made the

citations to be counted, is a philosophic as much as an empirical question. It is certainly

a role of librarians to enable researchers to optimize their awareness of the relevant

literature prior to publication, but no reasonable quantitative measures of such

optimization are conceivable. However, since peer review allows to be published only

those articles displaying an adequate awareness of the relevant literature, then perhaps

condition (ii) might be allowed, although the point is arguable. It might be claimed that

condition (iii), rational user behaviour, is commonly unmet because what researchers

tend to use is conditioned by what is most easily and quickly available to them at the

time. This point is taken further in the conclusion to this paper. At this point, while

acknowledging an issue here, it is assumed that the demands of the academic community

ensure that researchers behave rationally according to their perceived needs. There has

been much debate around condition (iv). Certainly, several analysts [17, 18, 19] have

used local citation analysis in combination with other techniques to estimate the actual or

potential use of journals, without reporting that the citation analysis produced anomalous

results. Comparing reshelving studies, professional judgement and local citation studies,

Swigger and Wilkes [20] note that "the citation method is the most reliable of the three".

More generally, Kelland and Young [21, 22] have reviewed the literature on relating

citation frequency with library use and concluded that the two are indeed related. Apart

from this empirical evidence, Kelland and Young also note that citation networks can be

considered as the collective memory of academic research. Since the selection decisions

undertaken by librarians contribute to both the access and archive functions of the library,

this feature of citations is also relevant. There is, of course, a significant literature critically analysing the motives and behaviour of citers [23, 24] but few such critiques significantly undermine the general assertion that those journals in which more highly cited articles occur are more suitable for consideration by a library manager than those in which less cited articles occur, especially when the citations in question come from local researchers. For example, MacRoberts and MacRoberts [25] note that citations only reflect the formally rhetorical aspects of science, which are not typical of the research process and of the majority of influences on any researcher. However, although librarians seek to support researchers in as many of their information requirements as possible, they (librarians) are principally interested in exactly that formal scholarly communication that the MacRoberts equate with citations. The final condition, that past use predicts future use, is one that must be met by all retrospective methodologies, including surveys, circulation and reshelving counts, and questionnaires. Whether or not it is met depends on too many factors to be adequately discussed in this paper.

One other charge that might be laid against the analysis of locally generated citations is that of impracticality. However, the work undertaken in generating and analysing such machine-readable citation counts is, in the experience of the authors, comparable with that of undertaking use surveys, and has the non-trivial advantage of not inconveniencing those whose journal use is being surveyed. One alternative methodology, the combined analyses of loans, interlibrary loan (ILL) requests and photocopies made, would involve considerably greater efforts than citation analysis, pushing it beyond the reach of most

libraries. Analysis of only one of these three variables would be vulnerable to criticism, for example, ILL requests alone do not measure journal use at all, but only some probable use of journals that are not held by the local library, especially if the local collection is checked before the requests are made, as is common in UK academic libraries.

A final criticism of the use of journal citations as indicators of potential journal use, and one that is likely to become increasingly pertinant, is that the use of electronic pre-print servers and other electronic, non-journal sources calls into question the use of the journal as a unit of analysis. This is not a criticism of citation analysis per se, but of limiting the analysis to journal articles, as this study does. The justification for focusing on journals is a practical one. The provision of journal articles is among the more difficult roles of the library manager, with rising subscription costs, falling library income as a proportion of university revenues [26], and multiplying sources in a more or less competitive market. While journal articles (in whatever form) retain a major role in the dissemination of scholarly information in many subjects, as is likely in the short and medium term, then research into their provision remains valid.

The use of local citation counts as indicators of potential journal use would appear to be defensible on both theoretical and practical grounds. The practicalities of acquiring and using such data are discussed as a part of the empirical study, below.

Core journals

Thus far, it has been argued that local citation counts represent a valid, practical and reasonably reliable measure of researchers' potential use of journals. If this is accepted, then the question arises as to how library managers can use journal title listings based on such counts to support selection and sourcing decisions. Clearly, the whole listing can be compared with the coverage of various sources, and that is done in the work reported below. However, most library managers would want to split their provision according to whether the journal in question was more or less important to researchers. This has been one context in which the idea of a core of journals has been raised. Traditionally, librarians have sought to hold collections of the most important printed journals, and to obtain other material on an ad hoc basis through mechanisms such as ILL. This divide in sourcing options is being maintained as both print collections and ILL evolve into the hybrid library environment [27, 28] because the economic rationale behind the divide remains the distinction between subscription and pay-per-item financial regimes. Hence, although the options open to librarians within each regime are now much broader than they have been, the need to distinguish between journals likely to be used more or less by researchers (the need to select) remains.

The primary theoretical basis of the notion of a core set of journals is, of course, that of Bradford's Law, although Bradford himself was a practising librarian and expressed the eponymous relationship in practical terms. In its original form, Bradford's Law is concerned with the relationship between the number of articles in particular journals that

relate to a particular discipline, and states that most such articles occur in a few journals central to that discipline [29]. Empirical work since then has shown that some citation patterns conform to the Bradford scatter [30]. The most common way of illustrating this is as a bibliograph (Figure 1). The bibliograph is constructed by listing journals in descending order according to the number of citations to them from a sample of articles. Each journal then has a rank order in this list. The cumulative total of citations is then plotted against the logarithm of the journal ranking. From data such as these, a 'nuclear zone' can be identified of journals that are cited disproportionately highly (in the non-linear part of the graph). In terms of the traditional bibliograph shown in Figure 1, it corresponds to c.
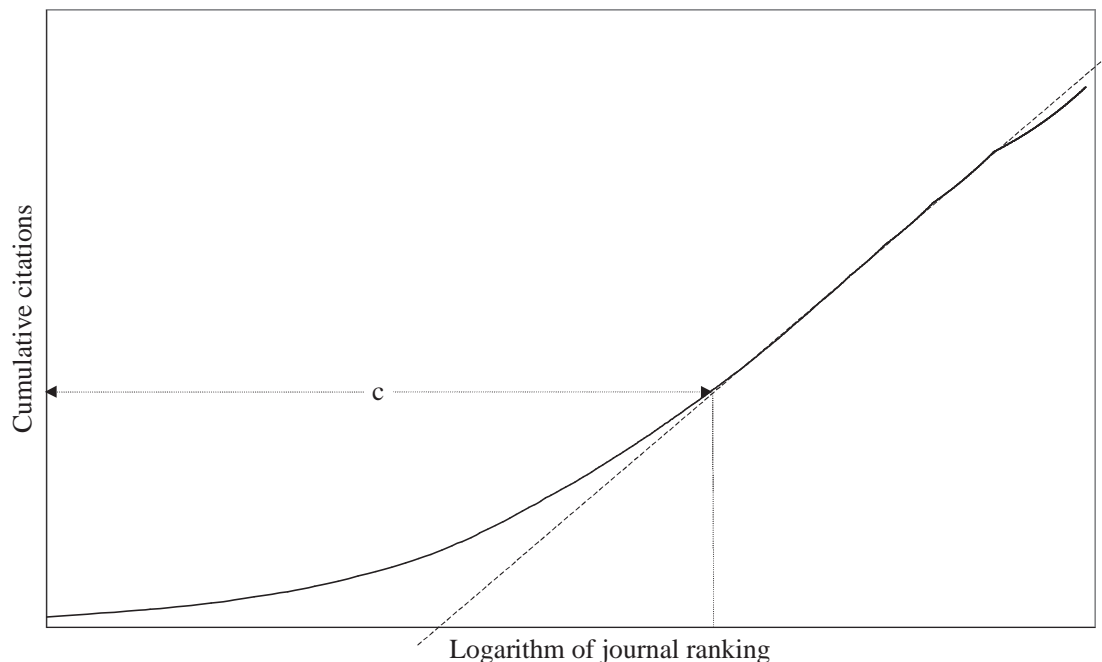

Figure 1. The Bradford distribution, showing the 'nuclear zone', c.

However, the definition of such a zone as a 'core' remains disputed [31, 32, 33]. For example, Rousseau and Leimkuhler [34] suggest that the core be defined as the region bounded by the origin and the point at which the gradient of the curve attains 75% of its maximum (linear) value.

Another theoretical approach to describing core journals is that employed by McCain [35] using co-citation data, in which related clusters or networks of journals are iteratively identified by the degree to which they are related by the citations contained therein. While being rigorous in itself, this approach is probably impractical for library managers.

In summary, despite theoretical underpinnings such as that of Bradford's Law, and despite librarians' need to identify core journals, there is no generally agreed, robust and practical definition of such a core. As White and McCain note:

"In practical work, cores are identified by investigators' purposes and can be expanded or contracted at will, no matter what someone's algorithm says." [36]

Nevertheless, the need remains to identify a proportion of highly cited journals against which the coverage of a number of sources can be compared. Several suggestions are apparent in the literature and were used in this study:

1. Bradford's 'nuclear zone', as defined by Brookes [37], which is the region of highly cited journals wherein the relationship between citations and the logarithm of the journal ranking is non-linear, see Figure 1.

2. The core defined by Rousseau and Leimkuhler [38] , wherein the gradient of the non-linear part of the Bradford curve is less than 75% of its maximum (linear) value.

3. The *n* most productive journals, where *n* is one more than one half of the number of 'singleton' journals, or journals cited only once, in the distribution. This is a definition derived from Goffman and Warren [39], and used by Gillaspy and Huber [40] and Brooks [41].

In addition, a fourth definition was used:

4. The smallest number of journals that can cover 80% of citations.

The fourth definition (a test of the Pareto 80/20 principle in this context) was included because the first three, especially the first two, concentrate on what many librarians might consider to be a restricted set of journals.

Each 'core' has a corresponding 'periphery' of journals that are cited, but insufficiently to be included in the core. These peripheral journals were included in the coverage analysis detailed below, alongside the core journals, since all journals used must somehow be obtained; as noted above, practising librarians work with a notion of 'core' in order to choose between subscription or pay-per-item financial regimes.

Journal coverage

Although potentially complex in practice, the conceptual basis of the coverage data was simple. Potential journal use, in terms of lists of journal titles developed from citation data, was compared with journal title lists from the sources used.

In the discussions above, mention has been made of the sources of journal articles. In the debate between access and holdings, the sources were document suppliers and local collections. In the hybrid library, the range of sources is now much expanded, to include online full-text databases, regional consortia and reciprocal arrangements, and linking technology that joins together the functions of secondary and primary publication and distribution. The sources assessed in this study are not limited to traditional document suppliers and local collections, but reflect this increased range of models. In this context, the old terminology becomes difficult to apply and it is necessary to think, not of suppliers, libraries and so on, but of *document access systems*. These are defined as combinations of social and technical arrangements necessary to get a document to a researcher's desktop in hardcopy (overwhelmingly preferred to electronic format for reading, according to FIDDO [42] and much other research [43]). In terms of traditional sources, one document access system is the set of printed journals held in a library, along with adequate photocopying equipment and a researcher willing to invest in the necessary work. Another document access system might be a desktop computer linked to an online database, itself linked to a document supplier, together with a postal service. Each document access system clearly needs financial and other management infrastructures.

The coverage of complex document access systems depends on the coverage of each of their elements (databases, suppliers and so on). Two types configuration are possible for multi-element document access systems, parallel and series. If a system has n elements in parallel then, in principle, users have access to the combined holdings of all n elements. An example of such a system in this study was *Combined holdings* (see Table 1), which was defined as the consolidated holdings of four university libraries. In cases such as these then, if the system's total coverage is denoted by C_p , and the coverage of each element denoted by c_1, c_2, and so on:

$$C_p = c_1 \cup c_2 \cup c_3 \dots \cup c_n$$

In contrast, if a system with n elements in sequence (series) has a total coverage denoted by C_s , and the coverage of each element is denoted by by c_1, c_2, and so on, then:

$$C_s = c_1 \cap c_2 \cap c_3 \dots \cap c_n$$

This is because the total coverage of the system is only those titles covered by all of its elements. An example of a system in this study including both types of configuration was *Geo-SilverLinker-BLDSC* (see Table 1), wherein users accessed in turn two combined databases (Geobase and Georef, searched together using SilverPlatter's ERL software) and then a document supplier (BLDSC).

As with the citation data, the practicalities of acquiring and using these lists are discussed as a part of the empirical study, below.

Comparing potential journal use with the journal coverage of systems

Scope of the study

The study took place against the backdrop of the work presented in the UK Anderson Report [44]. This raised the prospect of greater inter-institutional and regional collaboration in terms of UK scholars' access to resources. The study formed a part of the eLib FIDDO project [45], which found that nearly 40% of 113 researchers interviewed regularly used libraries other than that of their host institution, and that these were principally in nearby university libraries. Crotteau [46] has confirmed that visits to other libraries form a significant part of researchers' information gathering routines. Despite the suspicions of some librarians [47] that researchers are unwilling to use other libraries, it was felt that a regional focus offered the chance to assess user requirements and library provision at a valid level. Hence, it was decided to focus the study on a regional basis, specifically in four of the universities that are located within the UK Midlands area.

In order to carry out comparisons between potential journal use (as indicated by local citation counts) and the coverage of various sources, it was necessary to select a number

of subject areas. This was not only because several of the sources were subject-specific, but also because different subjects may have different citation patterns [48]. The regional basis of the study limited the choice of subject areas to those undertaken at most or all of the participating universities. A second criterion for the choice of subjects was that a substantial proportion of their formal scholarly communication should be in the form of journal articles. The subjects that best fit the criteria were business studies, geography and manufacturing engineering. It might be argued that researchers in some of these subjects rely significantly on non-scholarly material. Although this is likely to be true, it is irrelevant to the study of their potential journal use.

In the study described here, four general document access systems and three subject-specific document access systems were selected; they are described in Table 1 as they existed in 1998-9 when the fieldwork was conducted. These systems were selected because they covered a range of options that a librarian might consider. For example, the list includes both currently popular options within the UK university sector (for example, the British Library Document Supply Centre, *BLDSC*) and some that are, as yet, less popular. It also includes both general and subject-specific systems; an important point when interdisciplinary subjects are a focus of the analysis.

Table 1. The document access systems selected

Name	Subject coverage	Description
Combined holdings	General	The combined journal holdings of the libraries of the four universities included in this study.
BLDSC	General	The standard service offered by the British Library Document Supply Service via university ILL departments, supplying paper copies by a physical delivery system [49].
BL inside	General	A file of the 20,000 journal titles most requested from the BLDSC, and available to order by fax via a WWW interface [50].
ProQuest Direct	General	A desktop search and retrieval system available in 1998 from UMI (now Bell and Howell Learning Information), offering access to "one of the world's largest collections of information, including summaries of articles from over 8,000 publications, with many in full text, full image format." [51] Documents available to download in text / graphic or PDF formats.
SearchBank	Business	A system, similar to ProQuest Direct, that was available from the Information Access Corporation (Gale Group) during 1998, offering online access to business and marketing sources. Documents available to download in text / graphic or PDF formats [52].
Geo-SilverLinker-BLDSC	Geography	An example of linked systems, based on SilverPlatter's "SilverLinker" technology. The front-end databases were Geobase and Georef, searchable using the ERL software, and linked to BLDSC as a supplier [53].
EiText	Engineering	In this instance EiText (from Elsevier Engineering Information) was selected as a standalone supplier, although it is more often linked with the Compendex database. Documents were supplied as PDF email attachments [54].

Although not strictly relevant to this study, one point of interest might be the degree of

overlap between the four university library collections used as *Combined holdings.* The

numbers of journal titles common to the collections is shown in Table 2. The proportions

of each library journal collection that were unique to that collection are shown in Table 3.

Key to Tables 2 and 3:

Current subscriptions of university A: A

Current subscriptions of university B: B

Current subscriptions of university C: C

Current subscriptions of university D: D

Table 2. Coverage of the university collections making up *Combined holdings***

Definition	Number of journal titles	Percent of all titles covered by the four collections $(A \cup B \cup C \cup D)$
A	2,278	18.6%
B	4,799	39.3%
C	4,024	32.9%
D	4,927	40.3%
A∩B	487	4.0%
A∩C	492	4.0%
A∩D	488	4.0%
B∩C	1,066	8.7%
B∩D	921	7.5%
C∩D	1,151	9.4%
A∩B∩C	270	2.2%
A∩B∩D	280	2.3%
A∩C∩D	272	2.2%
B∩C∩D	603	4.9%
A∩B∩C∩D	198	1.6%
A∪B	6,590	53.9%
A∪C	5,810	47.5%
A∪D	6,717	54.9%
B∪C	7,757	63.4%
B∪D	8,805	72.0%
C∪D	7,800	63.8%
A∪B∪C	8,937	73.1%
A∪B∪D	9,983	81.7%
A∪C∪D	9,229	75.5%
B∪C∪D	10,912	89.2%
A∪B∪C∪D	12,226	100% (by definition)

Table 3. Proportion of each university's holdings (journal titles) that are unique to that collection

Definition	Number of journal titles	Proportion of each university's holdings
A-(B∪C∪D)	1,569	68.9%
B-(A∪C∪D)	3,671	76.5%
C-(A∪B∪D)	2,653	65.9%
D-(A∪B∪C)	3,458	70.2%

It is clear from Tables 2 and 3 that the overlap between the holdings of the four universities is relatively low. Table 3 shows that most of the titles appearing on each university's holdings list were unique to that university with the four universities of the sample. Because the figures above were calculated by matching journal titles, there were inevitably some errors due to inconsistent cataloguing practices between the four universities (a point taken further below). This would tend to understate the overlap between them. Nevertheless, the degree to which the local holdings of each institution are specific to that institution is striking.

The data

As noted in the introduction, the data were divided into two:

- citation lists of journal titles representing the potential use of journals by local researchers;

- coverage lists representing the net coverage of each document access system.

In order to obtain the citation data, lists were obtained of all active academic researchers within the selected subject areas in the four universities. From these lists, the ISI was able to produce Local Journal Utilization Reports, listing the journals cited by the named researchers in articles published by them between 1992 and 1997. These were appropriately combined to produce overall citation lists for the three subject areas selected; business studies, geography and manufacturing engineering. The generation of journal lists representing the four 'cores' (1, 2, 3 and 4, defined above) and 'peripheries' was undertaken from these overall lists.

The citation list for business studies contained 1,585 unique titles that were cited a total of 6,621 times, that for geography (from three universities) contained 1,234 unique titles that were cited a total of 5,971 times, and that for manufacturing engineering contained 1,293 unique titles that were cited a total of 5,374 times.

The coverage data were obtained from the various relevant bodies in the form of current subscriptions listings. No attempt was made to assess the length of backruns represented by any of the journal titles listed, as this would have made the survey impossibly ambitious. Instead, the current subscriptions listings were taken as representations of intended coverage. It is recognized that this could be seen as a drawback of the study reported here, unless reference is made to the second objective of this paper; to assess the

potential coverage of document access systems. The clearest indicator of <u>potential</u> coverage is likely to be current subscriptions.

Where more than one element contributed to the coverage of a document access system (such as in the case of *Geo-SilverLinker-BLDSC*), then these listings were combined as indicated above. The resultant coverage listings were then converted to a standard form and matched against the various citation listings. The results are shown below for each of the three subject areas, business, geography and manufacturing engineering.

The results

The figures given in the tables below are the proportions of citations covered by journals listed in the subscription lists of each document access system. The labelling conventions used are:

- each subject has a two-letter abbreviation, so that business is BS, geography is GY and manufacturing engineering is ME;

- the various 'cores' defined above are labelled 1-4, so that core 1 corresponds to the Bradford nuclear zone [34], core 2 is that defined by Rousseau and Leimkuhler [35], core 3 is that defined by Goffman and Warren [36], and core 4 is that which covers 80% of citations;

- each core of journals has a corresponding periphery, which includes those journals cited in the sample, but insufficently to be in the core; they are denoted by '.

For example, the region outside geography's Bradford nuclear zone is denoted as GY1'.

Business Studies

The four definitions of 'core' gave the results shown in Table 4. The coverage of the

relevant document access systems is shown in Table 5.

Table 4. The business cores

Core no.	Definition	Top N journal titles included	Proportion of all journals cited	Number of citations covered	Proportion of citations covered
BS1	Bradford 'nuclear zone'	N = 92	5.8%	2,576	38.9%
BS2	Rousseau-Leimkuhler	N = 73	4.6%	2,292	34.6%
BS3	Goffman-Warren	N = 356	22.5%	4,549	68.7%
BS4	Pareto principle	N = 575	36.3%	5,296	80.0% (by definition)

Table 5. The percentage business coverage of the document access systems

System	All citations	BS1	BS1'	BS2	BS2'	BS3	BS3'	BS4	BS4'
Combined holdings	78	88	71	87	73	83	66	82	60
BLDSC	98	100	97	96	100	99	96	100	93
BL inside	79	76	81	74	82	79	81	80	79
ProQuest Direct	7	10	6	10	6	9	3	9	3
SearchBank	11	16	7	17	8	14	4	12	3

The overall pattern of the figures is, perhaps, unsurprising. The *BLDSC* is able to supply

virtually all core material, and over 90% of peripheral material, regardless of the

definition of the core. The combined holdings of the four participating libraries and the

BL inside list each cover approximately 80%, whereas the newer, commercial WWW-based systems are much less successful in terms of coverage.

Geography

The four definitions of 'core' gave the results shown in Table 6. The coverage of the relevant document access systems is shown in Table 7.

Table 6. The geography cores

Core no.	Definition	Top N journal titles included	Proportion of all journals cited	Number of citations covered	Proportion of citations covered
GY1	Bradford 'nuclear zone'	N = 65	5.3%	2,647	44.3%
GY2	Rousseau-Leimkuhler	N = 22	1.8%	1,583	26.5%
GY3	Goffman-Warren	N = 270	21.9%	4,310	72.2%
GY4	Pareto principle	N = 398	32.3%	4,777	80.0% (by definition)

Table 7. The percentage geography coverage of the document access systems

System	All citations	GY1	GY1'	GY2	GY2'	GY3	GY3'	GY4	GY4'
Combined holdings	77	82	72	77	77	84	59	82	56
BLDSC	90	85	85	72	97	91	89	91	88
BL inside	73	64	81	49	82	72	75	72	76
ProQuest Direct	1	1	1	0	1	1	1	1	2
Geo-SilverLinker-BLDSC	39	43	55	53	33	43	26	43	23

The overall pattern for geography is similar to that for business studies, although the figures are generally lower. It is not clear why this might be so. Comparing the different systems, the combined library holdings have a relatively high coverage, especially for some of the definitions of core materials wherein *BL inside* fares less well. Less comprehensive still is the composite *Geo-SilverLinker-BLDSC* system, which seems restricted in its coverage by its databases (*Geobase* and *Georef*) rather than its supplier (*BLDSC*) component. *ProQuest Direct* has insignificant coverage of cited geography journals. More surprising, perhaps, is the finding that the narrower cores, GY2 and GY1, have lower coverage by all systems (including *BLDSC*) than the wider ones. This might be explained by the imperfect method of matching cited journal titles with titles covered by each system, further discussed in the review of the methodology, below. However, this effect is not seen in the results for business studies or manufacturing engineering. This fact, along with the generally lower figures for geography as compared with the other two subjects, would suggest that there is something specifically anomalous about the geography results.

Manufacturing engineering

The four definitions of 'core' gave the results shown in Table 8. The coverage of the relevant document access systems is shown in Table 9.

Table 8. The manufacturing engineering cores

Core no.	Definition	Top N journal titles included	Proportion of all journals cited	Number of citations covered	Proportion of citations covered
ME1	Bradford 'nuclear zone'	N = 122	9.4%	2,627	48.9%
ME2	Rousseau-Leimkuhler	N = 40	3.1%	1,503	30.0%
ME3	Goffman-Warren	N = 288	22.3%	3,666	68.2%
ME4	Pareto principle	N = 476	36.8%	4,299	80.0% (by definition)

Table 9. The percentage manufacturing engineering coverage of the document access systems

System	All citations	ME1	ME1'	ME2	ME2'	ME3	ME3'	ME4	ME4'
Combined holdings	85	99	71	100	77	93	68	91	62
BLDSC	97	98	97	97	97	97	98	98	96
BL inside	79	73	85	71	82	77	84	78	83
ProQuest Direct	3	3	3	2	3	2	4	3	3
EiText	37	48	26	59	28	42	26	40	26

The picture for manufacturing engineering is similar to that for business studies. The

BLDSC is, again, the most comprehensive system and, again, the consolidated holdings

of the four libraries in the region are slightly ahead of *BL inside*. The subject-specific

system, in this case *EiText*, covers rather less than a half of citations, and does appear to

be focused on core materials. Although its figures are slightly better than for geography,

ProQuest Direct remains a marginal system for engineers in terms of its coverage of

scholarly journals.

Evaluation of the study

Discussion of results

The four different definitions of 'core' journals result in a wide variation in the proportion of all cited journals and citations covered, from under 5% of journals and around 30% of citations for Rousseau and Leimkuhler's definition to the 80% of citations and over 30% of journal titles represented by the 'Pareto core'. Despite its importance within the information science literature, the cores based on the Bradford distribution (the Bradford nuclear zone and the Rousseau-Leimkuhler core) are probably too small to be useful for practising librarians deciding between subscriptions and access based on pay-per-item charging.

There are few surprises in the findings relating to the coverage of the various document access systems discussed in this paper. The British Library is widely regarded as having one of the most comprehensive research collections in the world, and the findings of this study would generally confirm their reported overall satisfaction rates [55] of over 91% for *BLDSC*. It should also be noted that both *BLDSC* and *BL inside* maintain a high level of coverage for both core and peripheral titles (they aim to be comprehensive), whereas the other document access systems generally have higher coverage levels of core than peripheral titles; they are focused on the more highly cited journals.

The *BL inside* coverage is defined as those titles most frequently requested from *BLDSC*. Maurice Line has asserted [56] that such requests are a significantly better measure of actual use than are citations. The very success of *BL inside* in covering journals accounting for around 80% of citations would imply, however, that there is a high degree of overlap between citations and requests to *BLDSC*.

A finding with, perhaps, more potential consequences is that relating to the coverage of the combined holdings of the four university libraries. The overlap between these libraries was low in terms of journal titles. That this list of titles, which was only just over 60% of the length of the list representing *BL inside*, should cover a comparable proportion of local citations is testament to the value of having skilled local collection managers deciding on the local holdings. It also confirms the potential for alternative systems to the British Library, such as LAMDA [57] in the UK although, as Anderson has pointed out [58], such systems rely on libraries that become net contributors being suitably recompensed. Equally, on the figures reported in this study, they would rely on the continued existence of the British Library for around one fifth of user requirements but, because these requirements would be from the least used journals, they would represent perhaps 60% of the total journal titles.

The performance of the newer document access systems seems disappointing. Some of them do not claim to be exclusively, or even primarily, scholarly resources. Those that do, for example, *EiText* and the system including *Geobase/ref*, generally had greater

coverage. Their value as supplements (rather than as potential alternatives) to local

holdings has been investigated by FIDDO, and the figures confirm that there is a lot of

overlap between them and coverage of local collections.

Discussion of methods

Using local citations as indicators of potential journal use by researchers is controversial,

but a defence of the practice has been offered. However, the conduct of this study has

revealed at least one further major problem with it. Comparing citation lists with

coverage lists involved choosing a field on which to make the comparison. Possibilities

included ISSN and title. Unfortunately, an ISSN field was not available with all of the

coverage lists and so, since a consistent approach was required across all systems, the

title field was used. This meant that the procedure was reliant on the consistency of the

cataloguing practices of all the document access systems and ISI. Although great efforts

were made to standardize the lists, it was clear that inconsistencies remained between the

automated comparisons (using Microsoft Access) and control comparisons undertaken by

hand. Adjustments were made for these inconsistences estimating a correction factor for

each comparison, based on the difference between the controls and equivalent automated

comparisons. However, this approach still led to some anomalies, especially with regard

to the results in geography, as noted above. A more satisfactory approach would be

highly desirable. However, despite problems such as this, a comparison of the results

with known figures (such as *BLDSC*'s satisfaction rate) suggests that the methodology is robust.

In terms of practicality, the method has strengths and weaknesses for library managers. Data collection was relatively straightforward, since all that ISI required were lists of researchers' names, along with their institutional affiliation. However, the cost of the processing of these lists against the ISI citation indexes might deter some; this study required 14 such lists to be processed. This cost $12,500 at 1998 prices. Analysis of the citation data was straightforward using simple database software. The most time was spent in standardizing the various lists of journal titles from ISI, databases and suppliers so that they could be compared, and in estimating the correction factor noted above.

Conclusions

This study has investigated both methodological and substantive issues. In terms of method, local citation analysis has been found to be a valid indicator of the potential use of journals in an academic research environment. In particular, it enables the identification of a number of 'cores' of journal titles, some of which appear more useful than others to library managers. However, there remain significant problems with the approach, especially when citation listings are compared with the coverage of systems of full-text delivery. Standardized cataloguing practices (or in future, perhaps, the use of digital object identifiers) would alleviate some of these problems. In terms of substance, the impressive achievement of the British Library Document Supply Centre in covering the journals cited by a sample of UK academic researchers has been confirmed. The British Library *inside* service was found to be comparable with a consolidated list of the holdings of four university libraries in terms of covering locally cited journals. Both of these held journals accounting for the vast majority of citations. Other systems were less successful, although their strengths might lie elsewhere, for example in terms of instant, desktop access to electronic resources. These features relate back to one of the conditions described toward the start of this paper for citations to be a valid indicator of potential use. Condition (iii), rational behaviour by researchers, assumes that the ease of use of particular sources of journal articles does not affect researchers' decisions about which journals to use. Clearly, the advent of instant, desktop access to some journals may have a substantial bearing on this assumption. As well as investigating their

coverage, the UK eLib FIDDO project has also conducted research into these other aspects of document access systems [59, 60].

1.   The FIDDO homepage is:
       http://www.lboro.ac.uk/departments/dis/fiddo/fiddo.html (visited 13 January 2000)

2.   Line, M.B. Draft definitions: information and library needs, wants, demands and uses. Aslib Proceedings, 26 (2), 1974, 87.

3.   Woolgar, S. Beyond the citation debate: towards a sociology of measurement technologies and their use in science policy. Science and Public Policy, 18 (5), 1991, 319-326.

4.   Line, M.B. Use of citation data for periodicals control in libraries: a response to Broadus. College and Research Libraries, 46 (1), 1985, 36-37.

5.   Horwill, C. and Lambert, P. 1 man - 100 votes: a new approach to reviewing periodicals subscriptions at the University of Sussex. Aslib Proceedings, 39 (1), 1987, 7-16.

6.   Greene, R.J. Computer analysis of local citation information in collection management. Collection Management, 17 (4), 1993, 11-24.

7.   Hurd, J.M., Blecic, D.D. and Vishwanathan, R. Information use by molecular biologists: implications for library collections and services. College and Research Libraries, 60 (1), 1999, 31-43.

8.   Broadus, R.N. A proposed method for eliminating titles from periodical subscription lists. College and Research Libraries, 46 (1), 1985, 30-35.

9.   Hoadley, I.B. Access vs ownership: myth or reality. Library Acquisitions: Practice and Theory, 17, 1993, 191-195.

10.   Brin, B. and Cochran, E. Access and ownership in the academic environment: one library's progress report. Journal of Academic Librarianship, 20 (4), 1994, 207-212.

11.   Line, M.B. Changes in rank lists of serials over time: interlending versus citation data. College and Research Libraries, 46 (1), 1985, 77-80.

12.   Zipp, L.S. Core serial titles in an interdisciplinary field: the case of environmental geology. Library Resources and Technical Services, 43 (1), 1999, 28-36.

13.   Johnson, W.T. Environmental impact: a preliminary citation analysis for local faculty in a new academic program in environmental and human health applied to collection development at Texas Tech University Library. Library and Information Science Research, 9 (1), 1999. Available at:
       http://aztec.lib.utk.edu/libres/libre9n1/toxcite.htm (visited 13 January 2000)

14. Walcott, R. Local citation studies: a shortcut to local knowledge. Science and Technology Libraries, 14 (3), 1994, 1-14.

15. Kreider, J. The correlation of local citation data with citation data from Journal Citation Reports. Library Resources and Technical Services, 43 (2), 1999, 67-77.

16. Line, M.B. Use of citation data for periodicals control in libraries: a response to Broadus. College and Research Libraries, 46 (1), 1985, 36-37.

17. Dole, W.V. and Chang, S.S. Survey and analysis of demand for journals at the State University of New York at Stony Brook. Library Acquisitions: Practice and Theory, 20 (1), 1996, 23-38.

18. Hughes, J. Use of faculty publication lists and ISI citation data to identify a core list of journals with local importance. Library Acquisitions: Practice and Theory, 19 (4), 1995, 403-413.

19. Schmidt, D., Davis, E.B. and Jahr, R. Biology journal use at an academic library: a comparison of use studies. Serials Review, 20 (2), 1994, 45-64.

20. Swigger, K. and Wilkes, A. The use of citation data to evaluate serials subscriptions in an academic library. Serials Review, 17 (2), 1991, 41-46.

21. Kelland, J.L. and Young, A.P. Citation as a form of library use. Collection Management, 19 (1/2), 1994, 81-100.

22. Kelland, J.L and Young, A.P. Citation patterns and library use. In: Kent, A. ed. Encyclopaedia of Library and Information Science. Volume 61, 1998. New York: Marcel Dekker.

23. Osareh, F. Bibliometrics, citation analysis and co-citation analysis: a review of literature 1. Libri, 46, 1996, 149-158.

24. Cronin, B. The citation process. London: Taylor Graham, 1984.

25. MacRoberts, M.H. and MacRoberts, B.R. Problems of citation analysis. Scientometrics, 36 (3), 1996, 435-444.

26. Campbell, R. The publisher's position: survival in an electronic world. In: Cunningham, A.M. and Wicks, W. (eds) Changing roles in information distribution. Philadelphia, PA: National Federation of Abstracting and Information Services, 1993.

27. Pinfield, S. et al. Realizing the hybrid library. D-Lib Magazine, October 1998. Available at: http://mirrored.ukoln.ac.uk/lis-journals/dlib/dlib/dlib/october98/10pinfield.html (visited 13 January 2000)

28. Oppenheim, C. and Smithson, D. What is the hybrid library? Journal of Information Science. 25 (2), 1999, 97-112.

29. Brookes, B.C. Bradford's Law and the bibliography of science. Nature, 224, 1969, 953-956.

30. McCain, K.W. Patterns of journal use in a departmental library: a citation analysis. Journal of the American Society for Information Science. 32 (4) 1981, 257-267.

31. Udofia. U.I. Selecting veterinary medical periodicals through citation analysis. Library Review, 46 (2), 1997, 105-112.

32. Depew, J.N. and Basu, S. The application of Bradford's Law in selecting periodicals on conservation and preservation of library materials. Collection Management, 8 (1), 1986, 55-64.

33. White, H.D. and McCain, K.W. Bibliometrics. In: Annual Review of Information Science and Technology, 24, 1989, 119-186.

34. Rousseau, R. The nuclear zone of a Leimkuhler curve. Journal of Documentation, 43 (4), 1987, 322-333.

35. McCain, K.W. Core journal networks and cocitation maps: new bibliometric tools for serials research and management. Library Quarterly, 61 (3), 1991, 311-336.

36. White, H.D. and McCain, K.W. Bibliometrics. In: Annual Review of Information Science and Technology, 24, 1989, 119-186.

37. Brookes, B.C. Bradford's Law and the bibliography of science. Nature, 224, 1969, 953-956.

38. Rousseau, R. and Leimkuhler, F.F. The nuclear zone of a Leimkuhler curve. Journal of Documentation, 43 (4), 1987, 322-333.

39. Goffman, W and Warren, K.S. Dispersion of papers among journals based on a mathematical anlaysis of two diverse medical literatures. Nature 221, 1969, 1205-1207.

40.   Gillaspy, M.L. and Huber, J.T. The literature of women and the Acquired Immunodeficiency Syndrome (AIDS): implications for collection development and information retrieval. Medical Reference Services Quarterly, 15 (4), 1996, 21-39.

41.   Brooks, T.A. Core journals of the rapidly changing research front of 'superconductivity'. In Borgman, C. (ed). Scholarly communication and bibliometrics. Newbury Park, CA: Sage Publications, 1990.

42.   FIDDO. Document delivery in a dynamic hybrid environment. Final report to the Electronic Libraries Programme. 1999.

43.   SuperJournal. Summary of SuperJournal findings. 1999. Available at: http://www.superjournal.ac.uk/sj (visited 13 January 2000)

44.   Anderson, M. Report of the Group on National/Regional Strategy for Library Provision for Researchers, report to Funding Councils (the Anderson Report). 1995. Available at: http://www.ukoln.ac.uk/services/elib/papers/other/anderson/ (visited 13 January 2000)

45.   See the FIDDO homepage, at: http://www.lboro.ac.uk/departments/dis/fiddo/fiddo.html (visited 13 January 2000)

46.   Crotteau, M. Support for biological research by an academic library: a journal citation study. Science and Technology Libraries, 17 (1), 1997, 67-86.

47.   Pat Wressel and Associates Library service provision for researchers. Proceedings of the Anderson Report Seminar organized by the Library and Information Co-operation Council and the Standing Conference of National and University Libraries. Bruton: LINC, 1997.

48.   Stankus, T. and Rice, B. Handle with care: use and citation data for science journal management. Collection Management, 4 (1/2), 1982, 95-110.

49.   BLDSC services are described at: http://www.bl.uk/services/bsds/dsc/ (visited 13 January 2000)

50.   The BL inside service is described at: http://www.bl.uk/online/inside/ (visited 13 January 2000)

51.   ProQuest Direct, now ProQuest, is described at: http://www.umi.com/proquest/ (visited 13 January 2000)

52. SearchBank, now InfoTrac, is described at:
    http://www.galegroup.com (visited 13 January 2000)

53. SilverLinker is described at the SilverPlatter site:
    http://www.silverplatter.com/ (visited 13 January 2000)

54. EiText is described at the Engineering Information site:
    http://www.ei.org/ (visited 13 January 2000)

55. British Library. Facts and figures 1999.
    http://www.bl.uk/services/bsds/dsc/pdf_files/factnfig.pdf (visited 13 January 2000)

56. Line, M. Changes in rank lists of serials over time: interlending versus citation
    data. College and Research Libraries, 46 (1), 1985, 77-80.

57. The Lamda service is described at:
    http://lamdaweb.mcc.ac.uk/ (visited 13 January 2000)

58. Anderson, M. Access to research collections in the UK: the Anderson Report
    updated. Library Review, 13 (4), 1998, 262-266.

59. Jacobs, N. et al. Planning document access: options and opportunities. London:
    Bowker Saur (forthcoming).

60. Morris, A., Woodfield, J. and Davies, J.E. Experimental evaluation of selected
    electronic document delivery systems. Journal of Library and Information Science,
    31 (3), 1999, 135-144.