Volume 35, Number 2 APRIL 1991 35(2):129-240 ISSN: 0024-2527

Yasar Tonta

Library Resources & Technical Services

ARTICLES

FEATURES

Association for Library Collections & Technical Services
a division of the American Library Association

A Study of Indexing Consistency between Library of Congress and British Library Catalogers

Yasar Tonta

Indexing consistency between Library of Congress (LC) and British Library (BL) catalogers using the Library of Congress Subject Headings (LCSH) *is compared. Eighty-two titles published in 1987 in the field of library and information science were identified for comparison, and for each title its LC subject headings, assigned by both LC and BL catalogers, were compared. By applying Hooper's "consistency of a pair" equation, the average indexing consistency value was calculated for the 82 titles. The average indexing consistency value between LC and BL catalogers is 16% for exact matches, and 36% for partial matches.*

For some time it has been observed that indexers tend to assign different index terms to the same document. In other words, "the indexers differ considerably in their judgment as to which terms reflect the contents of the document most adequately."[1] Essentially, indexing consistency is seen as "a measure of the similarity of reaction of different human beings processing the same information."[2]

Indexing consistency in a group of indexers is defined as "the degree of agreement in the representation of the essential information content of the document by certain sets of indexing terms selected individually and independently by each of the indexers in the group."[3]

Studies of indexing consistency reported in the literature have shown that the consistency values vary a great deal between indexers. Hooper, Leonard, and Markey reported the results of some 25 published and unpublished indexing consistency experiments in which the indexing consistency values ranged from 4% to 82%.[4,5,6] However, the indexing consistency scores of various studies, as researchers rightly caution us, should be considered separately and not compared. It appears that consistency values depend on a number of factors under which the indexing was performed. Zunde and Dexter listed 25 factors affecting indexing performance (see also Tarr and Borko[7,8]). For instance, factors such as the use of classification schedules and other indexing aids, the employment of subject specialists as indexers, and indexer training have greatly improved consistency values.[9,10] Markey offers a more detailed discussion, relating some of the factors to findings of previous studies.[11]

Yasar Tonta is a doctoral student, School of Library and Information Studies, University of California, Berkeley. Manuscript submitted March 25, 1990; revised October 4, 1990; accepted for publication October 17, 1990.

Another variable that affects indexing consistency is the consistency measure used in the evaluation. Studies reported in the literature employed a variety of methods and different formulae to calculate indexing consistency values. In fact, as Cooper puts it, "this circumstance makes generalization about their findings difficult."[12] (For more information about various indexing consistency formulae and statistical techniques involved in consistency studies, see Zunde and Dexter,[13] Hooper,[14] Leonard,[15] Markey,[16] and Rolling;[17] for a somewhat different method, see Chan.[18])

It is assumed that there is a relationship between indexing consistency and "indexing quality." That is to say, "an increase in consistency can be expected to cause an improvement in indexing quality."[19]

For some authors what is more important, and needs to be thoroughly scrutinized, is the relationship between indexing consistency and the effectiveness of information retrieval. Cooper further suggests that "until this relationship [i.e., the relationship between indexing consistency and retrieval performance] has been investigated, there is little point in measuring interindexer consistency at all."[20] Leonard attempted to investigate this relationship in his doctoral dissertation and found that "inter-indexer consistency and retrieval effectiveness exhibit a tendency toward a direct, positive relationship, i.e. high inter-indexer consistency in assignment of terms appears to be associated with a high retrieval effectiveness of the documents indexed."[21] However, he feels that "considerably more research is needed before the relationship between inter-indexer consistency and retrieval effectiveness can satisfactorily be defined."[22]

Methodology

This study represents an attempt to compare indexing consistency between Library of Congress (LC) and British Library (BL) catalogers.

For some time, BL catalogers ordinarily assigned LC Subject Headings to each document. The Bibliographic Services Division (BSD) of the BL was responsible for adding, among others, LC subject headings to UKMARC records. LC subject headings assigned by BL catalogers were based on the original analysis represented by the PRECIS (Preserved Context Index System) string and the corresponding *Dewey Decimal Classification* (*DDC*) number, not on subject analysis and consequent assignment of LC subject headings directly.[23,24]

In 1987, BL announced its final plans to change its subject cataloging policies starting in 1988. BL's three-stage cataloging plan envisioned, *inter alia*, "developing a specification for the future shape of PRECIS."[25] Since 1989 BL no longer assigns new LC subject headings to its British National Bibliography MARC (BNBMARC) records. Consequently, fewer UKMARC records contain LC subject headings.[26] Moreover, BL will replace PRECIS with a new subject indexing scheme in 1991. As a result of this change, LC subject data will no longer appear after this date in current BNBMARC records created initially by the BL.

As pointed out earlier, LC subject headings assigned by BL catalogers were based on PRECIS strings. Before assigning subject headings, BL catalogers did not check USMARC records to see if LC had already assigned subject headings to the titles in question. In other words, BL was not performing copy cataloging of LC subject headings data available in USMARC records.

Catalogers at LC do not check UKMARC records when performing subject cataloging either. At most they might see the British Cataloging-in-Publication (CIP) data on the verso of a title page during cataloging, but generally they disregard those subject headings. It has been LC's experience that subject headings assigned by the BL catalogers are not very useful.[27]

For the comparison of subject headings assigned by LC and BL catalogers, books published in the United Kingdom in 1987 in the field of Library and Infor-

mation Science (LIS) (020 in *Dewey Decimal Classification*) were chosen. First, all the titles published in 1987 were identified using the *BNB Subject Catalogue* (Vol. 1)—a total of 237. Using the ISBNs provided, all 237 titles were searched on the OCLC database. Of the 237 titles, records for 217 were found on OCLC. (The rest were either serials, microform copies, or local publications.)

Titles that were cataloged ("indexing" and "cataloging" are used interchangeably in this study) and given the *Library of Congress Subject Headings* (*LSCH*) by both LC and BL catalogers were identified. The 040 field in the MARC format was used to identify the origin of cataloging information. For instance, UKM stands for UKMARC, i.e., cataloged by BL, and DLC stands for LC, i.e., cataloged by LC. Items that were cataloged according to LC practices by libraries other than LC (by the National Library of Medicine, for example) are not included in the sample. By checking the 040 field for each record found on OCLC, it was possible to download all the records that were cataloged by both BL and LC. Eighty-two items were identified. Next, the LC subject headings assigned by BL and LC were compared for consistency.

For each item the headings found in fields 600 (personal name), 610 (corporate name), 611 (conference, congress, meeting, etc. name), 630 (uniform title), 650 (topical *LCSH*), and 651 (geographical *LCSH*) with second indicator value 0 (*LCSH*) were identified.

Finally, the "consistency of a pair of indexers" formula, defined by Rodgers and developed by Hooper, was applied to find out the indexing consistency value for each title cataloged by LC and BL catalogers. It was assumed that each individual cataloger at LC approaches the same document in the same way and assigns the same subject headings, which in fact might not be true. This assumption was also made for BL catalogers. In fact, what is found is *not* the individual inter-indexer consistency value between the two indexers but, rather, the indexing consistency value between LC and BL catalogers as two different groups.

According to Hooper's equation, "the consistency of one indexer with respect to a second is based on the number of times the two indexers agree on the use of a term, divided by the total number of terms used by either indexer (based on the specific document)."[28]

Hooper's "consistency of a pair" formula is as follows:

$$CP(\%) = \frac{A}{A + M + N}$$

where CP is the consistency of term assignment between two indexers (consistency expressed as a percentage); A is the number of term agreements between "M" and "N" for a specific document; M is the number of terms used by "M" but not used by "N"; and N is the number of terms used by "N" but not used by "M."

Having obtained the indexing consistency value for each title, the average indexing consistency value between BL and LC catalogers for the 82 titles was calculated.

Findings

The major findings of the study are as follows:

1. LC catalogers assigned 282 subject headings for 82 items while BL catalogers assigned 127 for the same 82 items. In other words, on the average, LC assigned 3.44 subject headings per title (SD = 1.97) whereas BL catalogers assigned 1.55 subject headings (SD = 0.79). There seems to be a weak association (r = 0.34) between the LC and BL catalogers as two different groups in terms of the number of subject headings assigned for each item. (In fact, the correlation coefficient goes down to 0.20 when an outlier is excluded from the calculation.)

   The marked difference between the average number of subject headings assigned by LC and BL catalogers is understandable. It is obvious that BL relies on PRECIS

for subject access rather than *LCSH*, whereas LC completely depends on *LCSH* for subject retrieval.

This weak association between the LC and BL catalogers as two different groups might, on the other hand, reflect differences in the experience and expertise of subject catalogers and the depth of subject indexing. During the subject analysis of a title, catalogers often identify some obvious concepts that later become "legitimate" subject headings. Some catalogers nevertheless overlook some of these concepts and therefore do not assign otherwise useful subject headings for particular titles. More importantly, LC and BL might have had somewhat differing policies regarding the depth of indexing, which would profoundly affect the number of headings assigned by their catalogers. No matter how competent the subject catalogers in each institution are in assigning subject headings, a strong association cannot be expected if, for example, due to economic considerations, one of the institutions limits the maximum number of subject headings per title regardless of the characteristics of the titles. Findings of the present study suggest that LC is more liberal than BL in assigning subject headings: LC assigned, on the average, 3.44 subject headings per title compared to BL's 1.55 subject headings per title.

It appears that BL catalogers tend to keep the number of headings assigned for each title to a minimum. Only for 2 titles (2.4%) did BL catalogers assign more subject headings than LC catalogers. BL and LC catalogers assigned the same number of subject headings for 17 titles (20.7%). It should be stressed, however, that assigning the same number of subject headings for each item does not necessarily mean that they assigned the same subject headings for each item. For the remaining 63 titles (73.9%) LC catalogers assigned more *LCSH* than BL catalogers.

2. Each and every subject heading for the same title that was assigned by LC and BL catalogers was compared. Forty-nine out of 127 BL-assigned subject headings *exactly* matched the LC-assigned subject headings. "Exact matches" included variants in spelling (i.e., catalog–catalogue) and punctuation (i.e., on-line–online), but not synonyms (i.e., non-book–audio-visual).

The following are examples of "exact matches":

a. Title: *Reference services today: from interview to burnout*
   LC: Reference services (Libraries)
   BL: Reference services (Libraries)
   (Both subject headings exactly match. Note that the example above is also a "perfect match," i.e., the indexing consistency is 100%.)

b. Title: *A guide to collecting librariana*
   LC: Library science—Collectibles
   Libraries—Collectibles
   Bibliography—Collectibles
   Book collecting
   BL: Libraries—Collectables
   (The second LC-assigned subject heading and BL's only heading match exactly except for spelling.)

The example below is *not* considered an "exact match." Although the second LC-assigned subject heading and the BL-assigned one are conceptually the same, synonyms were used (i.e., audio-visual—non-book); such subject headings were treated as "partial matches" in this study.

c. Title: *Legal deposit of non-book materials*
LC: Libraries—Special collections—Non-book materials
Acquisition of *non-book* materials
Acquisition of non-book materials—Great Britain
Legal deposit (of books, etc.)
Legal deposit (of books, etc.)—Great Britain
BL: Acquisition of *audio-visual* materials

By applying Hooper's equation for exact matches, the average indexing consistency value between BL and LC catalogers was found to be 16%. (Further examples of subject headings assigned by LC and BL catalogers for identical titles are given in appendix A.)

3. In the second run *partial* matches were added. Forty-four BL-assigned headings partially matched further. A synonym in a multiple-word-subject-heading was treated as a "match" as long as it was not the first word in that subject heading. The lack of a subdivision in a subject heading was also accepted as a partial match if the main part of the subject heading matched exactly. The following are examples of "partial matches":

a. Title: *Access to local authority official publications: proceedings of a seminar*
LC: Local government documents—Great Britain—Congresses
Local government documents—Information services—Great Britain—Congresses
BL: Local government documents—Great Britain—Bibliography—Methodology

b. Title: *Reference and information services: a reader for today.*
LC: Reference services (Libraries)
Information services
BL: Reference services (Libraries)—United States
Information services—United States

Note that although all subject headings above have the same main headings, subdivision(s) differ. Such headings were treated as "partial matches." It should also be noted that in large online catalogs the lack of a subdivision in a subject search will yield many irrelevant hits as well as relevant ones, thereby increasing the information "overload." Consider, for instance, the subject headings **Library science** vs. **Library science—Automation.**

The example below is *not* a "partial match" even though the first words in the first LC-assigned subject heading and the second BL-assigned one are the same.

c. Title: *Design and production of media presentations for libraries*
LC: Audio-visual *library service*
Library science—Audio-visual aids
Communication—Audio-visual aids
Media programs (Education)
BL: Library orientation—Aids and devices
Audio-visual *materials*

For both exact and partial matches, the average indexing consistency value between BL and LC catalogers was found to be 36%. (Several examples of consistency values are given in appendix A.)

Seventeen BL-assigned subject headings for 12 titles were completely different from those assigned by LC.

4. Assuming that the indexing consistency value between BL and LC catalogers would have been different if the number of subject headings assigned by BL catalogers were equal to that of LC catalogers, the indexing consistency value was calculated for 17 titles that have the same number of *LCSH* assigned by both LC and BL indexers. The following indexing consistency values were obtained:

   For example matches, the average consistency value was found to be 14%.

   For both exact and partial matches, the average consistency value was found to be 41%.

   Although there is a slight difference between the two averages, (i.e., 16% vs. 14% for exact matches, and 36% vs. 41% for partial and exact matches), there seems to be no strong relationship between the indexing consistency value and the assumption that if an equal number of subject headings were assigned by both LC and BL for all titles, the consistency value would have been different.

Conclusions

Findings obtained in this study suggest that the indexing consistency value between LC and BL catalogers for books in the field of Library and Information Science is rather low: 16% for exact matches and 36% for both exact and partial matches. In fact, these low indexing consistency values verify the findings of previous studies.

Low indexing consistency values between LC and BL catalogers might have some implications for copy cataloging. Copy cataloging of UKMARC records bearing *LCSH* could produce some surprising results for LC catalogers. Such surprises should also be expected by BL catalogers. It appears that LC and BL catalogers use somewhat varying terminologies, at least in Library and Information Science; they often disagree on which indexing terms to assign for a particular title. Using somewhat different (or, at least not the same) index terms may well be due to the fact that both LC and BL catalog materials according to the requirements of their clients and/or users. It could be that certain terms are not commonly used on both sides of the Atlantic. Nevertheless, indexing consistency rates should be taken into account when performing copy cataloging.

The fact that BL assigns fewer subject headings than LC has some important consequences regarding subject access in library catalogs. Assigning more subject headings per title increases the number of subject access points for a given title. It is reasonable to suggest that titles posted under various subject headings will be more accessible, though not necessarily more useful, than those posted under fewer subject headings.

Although it is difficult, or, indeed, inconceivable, to extend the findings obtained in this study to other fields, the following can be said of indexing consistency in general terms:

Indexing consistency is certainly an important issue and should be studied further. Similar studies comparing more titles in other fields as well as in Library and Information Science could be conducted.

It seems that much remains to be done to improve indexing consistency between professional indexers, even though controlled vocabularies such as *LCSH* are helpful. No matter how competent and experienced the indexers are, there is no guarantee that using the same tools, at least in phrasing subject headings, will ensure consistency among different indexers in assigning topical subject headings.

It is widely believed that catalog users have some understanding of current subject headings, at least in their respective fields, so that, by using the subject approach, they can retrieve what they want. The findings of indexing consistency studies, however, do not support this commonly held view. From the users' point of view, the more consistent the indexing terms are, the less frustrated the users get when searching cata-

logs. After all, one would not expect users to guess the "right" subject headings correctly all the time if indexing consistency were low.

One should also consider the consistency (or, rather, variety) in users' vocabulary/terminology when naming the same concepts. This has a profound effect on the overall success in searching library catalogs. In fact, researchers have found that considerable numbers of subject searches in online catalogs resulted in no retrievals due to, among other factors, lack of knowledge concerning *LCSH* terminology and misspellings.[29] Percentages of zero retrievals in subject searching range from a low of 35% to a high of 57.5%.[30]

Perhaps more important than the inter-indexer consistency is the consistency between the terminology of indexers and that of catalog users. At present, most of the controlled vocabularies provide limited numbers of cross-references in order to refer the user to the preferred indexing terms. The development of online catalogs with subject searching facilities will enable us to study the consistency issue further. For example, the availability of *LCSH* online in online catalogs makes it possible to compare the users' vocabulary with *LCSH* terminology and to see how much discrepancy exists between the two. If the users keep entering the same index terms for a particular subject and those index terms are not available in the system as "legitimate" subject headings, the terms could be changed or new cross-references could be added. Such experiments would certainly add a new dimension to indexing consistency studies and improve the success rate in subject searching in online catalogs.

References and Notes

1. Pranas Zunde and Margaret E. Dexter, "Indexing Consistency and Quality," *American Documentation* 20, no.3: 259–67 (1969).
2. Zunde and Dexter, "Indexing Consistency," p.259.
3. Ibid.
4. R. S. Hooper, *Indexer Consistency Tests—Origin, Measurements, Results and Utilization* (Bethesda, Md.: IBM Corp., 1965).
5. Lawrence E. Leonard, *Inter-indexer Consistency Studies, 1954–1975: a Review of the Literature and Summary of Study Results*, University of Illinois Graduate School of Library Science Occasional Papers no.131 (Urbana, Ill: University of Illinois, 1977).
6. Karen Markey, "Interindexer Consistency Tests: a Literature Review and Report of a Test of Consistency in Indexing Visual Materials," *Library and Information Science Research* 6, no.2:155–77 (1984).
7. Pranas Zunde and Margaret E. Dexter, "Factors Affecting Indexing Performance," in *Proceedings of the American Society for Information Science*, vol. 6, 32nd Annual Meeting, ed. Jeanne B. Nash (Washington, D.C.: ASIS, 1969), p.313–22.
8. Daniel Tarr and Harold Borko, "Factors Influencing Inter-indexer Consistency," in *Proceedings of the American Society for Information Science*, vol. 11, 37th Annual Meeting, ed. Pranas Zunde (Washington, D.C.: ASIS, 1974), p.50–55.
9. Hooper, "Indexer Consistency Tests," p.7.
10. Zunde and Dexter, "Indexing Consistency," p.260.
11. Markey, "Interindexer Consistency Tests," p.155–77.
12. William S. Cooper, "Is Interindexer Consistency a Hobgoblin?," *American Documentation* 20, no.3:268–78 (1969), p.269.
13. Zunde and Dexter, "Indexing Consistency," p.259–67.
14. Hooper, "Indexer Consistency Tests."
15. Leonard, "Inter-indexer Consistency Studies."
16. Markey, "Interindexer Consistency Tests," p.155–77.
17. L. Rolling, "Indexing Consistency, Quality and Efficiency," *Information Processing & Management* 17, no.2:69–76 (1981).
18. Lois Mai Chan, "Inter-Indexer Consistency in Subject Cataloging," *Information Technology and Libraries* 8: 349–58 (1989).
19. Cooper, "Is Interindexer Consistency a Hobgoblin?" p.269.
20. Ibid.
21. Leonard, "Inter-indexer Consistency Studies," p.33.
22. Ibid.

23. Derek Austin, *PRECIS: a Manual of Concept Analysis and Subject Indexing.* (London: British Library, Bibliographic Services Division, 1984), p.286.
24. For an earlier account of the use of LC data in the British Library, see John Gilbert, "Library of Congress Data in BSD," *British Library Bibliographic Services Division Newsletter* no.11 (1978) p.2–3.
25. For a more detailed discussion of the British Library's subject indexing proposals, see "Currency with Coverage," *British Library Bibliographic Services Newsletter* no.44 (1987) p.1–3, and "Currency with Coverage; Subject Indexing Proposals," *British Library Bibliographic Services Newsletter* no.45 (1988) p.1–3.
26. Personal communication of one of the referees, April 9, 1990. The author would like to thank the editor for providing the transcripts of the referee's communication with the British Library and OCLC.
27. Mary K. D. Pietris, personal communication, July 31, 1990.
28. Hooper, "Indexer Consistency Tests," p.33.
29. Ray R. Larson, "Managing Information Overload in Online Catalog Subject Searching," in *ASIS '89: Proceedings of the 52nd ASIS Annual Meeting Washington, D.C., October 30–November 2, 1989,* ed. Jeffrey Katzer and Gregory B. Newby (Medford, N.J.: Learned Information, Inc., 1989), p.129–35.
30. Karen Markey, *Subject Searching in Library Catalogs: Before and After the Introduction of Online Catalogs.* (Dublin, Ohio: OCLC, 1984).

Appendix A: Examples of Indexing Consistency Values between Library of Congress and British Library Catalogers

1. Vickery, B. C. and Vickery, A. *Information science in theory and practice*
   LC: Information science
   BL: Information Science
   Consistency value: 100%
2. Veit, Fritz. *Presidential libraries and collections*
   LC: Presidents—United States—Archives
   BL: Presidents—United States—Archives
   Consistency value: 100%
3. ur Rahman, Sajjad. *Management theory and library education*
   LC: Library education
   Library administration—Study and teaching
   Library administrators—Training of
   BL: Library administration—Study and teaching
   Library education
   Consistency value: 67%
4. *Personnel issues in reference services*
   LC: Reference librarians
   Library personnel management
   Reference services (Libraries)
   Library administration
   Library services—Organization & administration
   BL: Library personnel management
   Reference services (Libraries)
   Reference Librarians
   Consistency value: 60%
5. Crawford, Walt. *Technical standards: an introduction for librarians*
   LC: Library science—Technological innovations—Standards
   Library science—Standards
   Technology—Standards
   Information science—Standards
   BL: Library science—Standards
   Information science—Standards
   Consistency value: 50%
6. Strickland-Hodge, Barry. *How to use Index Medicus and Excerpta Medica*
   LC: Medicine—Abstracting and indexing
   Index medicus
   Excerpta medica
   Medicine—Bibliography—Methodology
   MEDLARS-MEDLINE information system—United States
   Abstracting and Indexing
   BL: Index Medicus
   Excerpta Medica
   Medicine—Abstracting and indexing
   Consistency value: 50%
7. Burton, Paul F. *The librarian's guide to microcomputers for information management*
   LC: Libraries—Automation
   Library science—Data processing
   Microcomputers—Library applications
   BL: Microcomputers—Library applications
   Consistency value: 33%
8. Harrod, Leonard Montague. *Harrod's librarians' glossary of terms used in li-*

brarianship, documentation, and the book crafts and reference book
LC: Library science—Dictionaries
Information science—Dictionaries
Bibliography—Dictionaries
Book industries and trade—Dictionaries
BL: Library science—Dictionaries
Consistency value: 25%

9. *Conservation of library and archive materials and the graphic arts*
LC: Library materials—Conservation and restoration
Archival materials—Conservation and restoration
Graphic arts—Conservation and Restoration
Books—Conservation and Restoration
Art—Conservation and Restoration
Paper—Preservation
BL: Library materials—Conservation and restoration
Consistency value: 17%

10. Tracy, Joan I. *Library automation for library technicians*
LC: Libraries—Automation
Library science—Data processing
Library technicians
BL: Processing (Libraries)—United States
Consistency value: 0%

11. Hartley, J., Noonan, A., and Metcalfe, S. *New electronic information services*
LC: Database industry—Great Britain
Database industry
BL: Information storage and retrieval systems
On-line data processing
Consistency value: 0%

12. *The Application of Microcomputers in Information, Documentation, and Libraries: Proceedings of the Second International Conference on the Application of Micro-Computers in Information, Documentation and Libraries*
LC: Libraries—Automation—Congresses
BL: Microcomputers—Library applications
Consistency value: 0%