THE RELIABILITY OF TWO CITATION CHECKING TECHNIQUE MEASUREMENT INSTRUMENTS

Silas Marques de Oliveira
PUCCAMP

RESUMO

OLIVEIRA, S.M. *Fidedignidade de duas técnicas de checagem de instrumentos de medida.* **Transinformação,** *4 (1,2,3): 55 - 79, 1992.*

A importância da aplicação de instrumentos de medidas fidedignos quando se avalia o valor da coleção de uma biblioteca é discutidamente, tudo. O teste-reteste de fidedignidade foi usado para determinar a fidedignidade de duas técnicas diferentes de medida de citação, especificamente, citações derivadas de periódicos e derivados de monografias. O chi-quadrado para igualdade de proporções demonstrou que os periódicos podem ser instrumentos mais fidedignos sendo levantadas questões quanto a fidedignidade das monografias.

Unitermos: *fidedignidade, avaliação de coleção, instrumentos de avaliação.*

1 - INTRODUCTION

One of the most serious problems that exists in the library is collection development. In 1977 BAUGHMAN (1977, p. 241) stated that during the 1960's, "the prevailing slogan 'the bigger the better', led one to believe that excellence relates directly to the quantity of volumes that the library holds". He continues, "Ironically this solgan has forged the librarian's action into the kiva of materials administration rather than meaningful collection development - the raison d'etre of the library."

However, continous assessment of the library's holdings is vital to an effective collection development plan. ALA's (1979) **Guidelines for Collection Development** state that only by means of an evaluative study can one

determine whether the collection is meeting its purpose, or is serving its users or even in which ways or areas it is deficient or strong. Evaluative methods are needed to gather information on the scope, quality, accessiblity and usefulness of the existing collection so that collection development can respond to the current needs of the library's users, avoid waste of resources, and unbalanced growth.

Al though the evaluation of library holdings is an essential managerial task, very few studies address the reliability of the instruments used to measure the collection's strengths and weaknesses. One commonly used measurement of collection strenghts is the "citation checking technique" (CCT). CCT attempts to assess collections through the comparison of citations in publishecd sources with the holdings of the library. Two assumptions of CCT are: (1) citatoins are an indicator of use; (2) the library's ownership of sources cited in published works or dissertations is an indicator of the quality of strength of the collection.

There have been numerous applications of CCT to collection evaluation: BLAND, 1980; CHAMBERS, 1973; COMER, 1981; BROADUS, 1971; MARTIN, 1952; SINEATH, 1970; STEWART, 1970; STRAYER, 1971, to mention only a few. However, Since there has not been a consistent effort to determine the reliability of the measuring instruments used in applying CCT, the purpose of this study is to undertake an analysis of two different CCT measurement instruments in an attempt to determine their reliability. A measurement instrument is characterized by the type of source from which the citations are gathered and by parameters of research design utilized.

2 - THE CONCEPT OF RELIABILITY

Reliability is seen as an indication of stability in research results. In order to have confidence in the results of a research study, it is necessary that the measuring instrument utilized in the investigation be reliable. The term "reliability" is most commonly used to characterize consistent, dependable, and stable research methods, instruments or results.

This study uses "reliability" as GOVE (1961, p. 1917) defines it: "the extent to which an experiment, test, or measuring procedure yields the same result on repeated trials." The findings of this study could have significance in the determination of which of the measurement instruments tested is more reliable for collection evaluation purposes. If either of these is found to be reliable, library managers will possess an adequate measurement instrument to accomplish their most important mission - bring together people and information in a meaningful way.

Several research methods specialists emphasize the replicability aspect of research as a means of testing the reliability of the instrument. KIDDER (1980), p. 7) observes that "generally speaking, research is considered to be... reliable when the findings are repeatable. Regarding the design, the researcher should ask whether the conclusions are... repeatable (reliable). Regarding the measurement process, the concern should be with whether the sources of observations are... repeatable (reliable). "For BUSHA & HARTER (1980, p. 97), "when studies are repeated, the reliability of previous research data can be tested."

Commenting on research designs, POWELL (1985, p. 37) also defends the notion of being able to replicate. He states that "reliable research findings are repeatable. That is to say, if a study is duplicated, or exactly repeated, utilizing the same procedures and techniques, the results should be the same. And, within limits, the findings of a study high in reliability can be generalized beyond the particualr conditions in the research, at least so long as the conditions remain essentially the same."

Analyzing the main concepts of these statements, one concludes that in order to test the reliability of a measurement instrument, it is necessary to test the results more than once. But the results should be retested in settings as similar as possible. A measurement is generally considered to be reliable when the error component is reasonably small and does not fluctuate greatly from one observation to another.

The "Test-Retest" reliability test was used in this study. Employing this technique, the researcher uses the same data collection instrument to observe or collect scores two or more times from the same group of subjects, under conditions which are judged to be very similar. The results of these measurements are compared to determine their similarity. KINNEAR & TAYLOR (1987) explain that this approach assumes that the greater the discrepancy in the results the greater the random error present in the measurement process and the lower the reliability.

3 - CHARACTERISTICS OF CCT

As noted earlier, CCT ia a process of developing a list of citations gathered form different sources and comparing it against the library's files to determine the percentage the library has in its own collection. Presumably, a high percentage of items found indicates successful collection development. In contrast to standard bibliographic lists which are compiled by experts, CCT is built based on the bibliographies, footnotes and references in primary sources such as books, journals and dissertations. Usually there are no selection criteria for the compilation of these lists from a set of source

documents. The items which form the lists are most frequently selected at random from the primary sources without any consideration for their contents. More often than not, the sources which contain the citations that make up the citation lists are chosen based on pre-established criteria, such as subject area, period of coverage, languages, etc.

These lists are intended to be used in a specific study, usually to evaluate the collection of a specific library. The specificity of purpose in compiling these lists is its asset when compared to other list-checking techniques. The fact that a list is generated randomly means that personal biases are less likely to enter into the selection of the citations. Since CCT is based on the principle that the actual use of material is indicative of its relevance to current research, the results of CCT studies are quantitative and empirically based data from which a qualitative judgment can be made about the ability of a collection to support research in a specified area. The CCT process is often used to determine whether or not a specific scholarly work could have been written with the resources of the library being evaluated. Thus, CCT presents a unique characteristic which distinguishes it from other evaluation methods - while it uses quantitative data it also permits a judgment to be made on the quality of the collection.

Several other advantages of CCT are mentioned in the library and information science literature:

(a) The citations gathered for the evaluation are limited to a restricted segment of the collection. COALE (1965) argues that the size of the collection poses a problem to most methods of evaluation, but CCT permits evaluation in selected subject areas within the collection.

(b) Some of the citations will be of a peripheral nature, which will allow the evaluator to verify if a library lacks materials on subjects tangential to a specific area. Commenting on this aspect of the CCT, COALE (1965, p. 174) mentions that a library "might have quite a good special collection and yet be a poor place for a scholar to work if many necessary titles tangential to his subjcet are lacking." Also LANCASTER (1977), COMER (1981), and other authors consider the collection stronger if it is able to offer tangential material to its users. There is a greater probability that a list developed using CCT will contain marginal materials because it is formed by selecting items at random.

(c) "The collection is evaluated not against some theoretical list of best books, but against lists of sources actually consulted by scholars writing in the field." This advantage offered by LANCASTER (1977, p. 176) is supported by other authors, such as PRICE (1963) and BONN (1974). PRICE (1963) comments that there appears to be an average conscientiousness in giving credit to papers that have provided the foundation for the work. This is in accordance with BONN (1974) who states that the fact that a source is actually

used to support research is per se an indication of its value or quality. It is also indicative of its relevance to current research, producing empirical data for a qualitative judgment about the ability of a collection to support research, thus being a useful item to be checked against the collection which is being evaluated.

(d) Specially compiled lists are much more effective than standardized, published lists, LANCASTER (1977), COMER (1981), BONN (1974) and MCINNIS (1983) all agree with this statement. BONN (1974, p. 275, 279) emphasizes this in two passages:

"Especially compiled lists that are tailored to the particular library or libraries and for well defined purposes are generally considered much more reliable as evaluators of quality than are the readily available published lists."

"For the most fruitful results the checklists used must be carefully selected or especially ccompiled to match the needs of the survey and the goals and objectives of the library or libraries being surveyed."

(e) CCT can easily be undertaken by a competent searcher or experienced cataloguer. Since this technique does not involve any type of content analysis, a responsible person with a good knowledge of sampling technique can undertake this evaluation process. There is no risk of content misinterpretation, and it allows for great economy of time and money.

Although the CCT is characterized by numerous advantages, it is not free from limitations. Several authors present shortcomings of this technique. BROADUS (1977a & 1977b) points out several disadvantages in the use of the CCT as a tool for collection evaluation purposes. A fundamental is that it shows what scholars have done, whereas the more important question is what they should have done; there is no way to tell whether authors really consult the materials which are best for their purposes; does not make distinctions between the best and the worst titles mentioned; the method would be difficult or impossible to use for evaluatin a general collection, because of the difficulty of constructing a multi-subject checklist from citation counts.

NISONGER (1983, p. 164, 168) also suggests several disadvantages of this technique: "The most recently published literature is not included or, at minimum, is underrepresented; abstracts and indexes are seldom cited; secondary sources are underrepresented; and the technique is oriented towards the needs of library patrons who publish."

Although aware of the problems and limitations that CCT can present, the library administrator should not ignore the positve and beneficial aspects of this method, MOSHER (1984, p. 214-215) states that "citation studies continue to surface as the most efficient and informative single form of evaluation study of research collections." VOOS (1981) considers citation

analysis a viable means, and perhaps, the best objective approach to collection evaluation.

The ideal tool for evaluating a collection, according to NISONGER (1983, p. 163) is a "scientifically based, versatile collection evaluation technique that results in empirical data and can be implemented at a relatively low cost to the library... the use of citation checking as an evaluation tool", he comments, "promeses to meet many of the above criteria." Thus, one is led to agree with NISONGER's (1980, p. 337) final remarks about this evaluation technique: The CCT actually "does measure the depth of the collection and, consequently, would constitute a valid evaluation tool." The fact that many authors have reported studies using this technique suggests that they share Nisonger's view.

4 - CHARACTERISTICS OF CCT STUDIES

Fourteen studies which applied the CCT to evaluate library collections or parts of them have been identified. The majority of these evaluations were conducted at academic/university libraries. Investigators have used different types of documents as sources of citations. The numbers of citations used in these studies also varies, from 250, BLAND (1980) to 7.000, CHAMBERS & HEALEY (1973). Al though it is evident that there is a tendency towards evaluating areas of the social sciences, the evaluative studies did not present any uniformity in relation to the subjcet areas of the collection which were evaluated. Except for the works of LOPEZ (1983) and NISONGER (1980), all of the CCT studies reported in the library and information science literature used a simple percentage count of matches between the list of citations and the library holdings to measure the strength of the collection.

These parameters for research design used in those studies are summarized in Table I. As the Table demonstrates, there is no uniformity in their use. However, we observe that most of those investigations did not question the reliability of the measurement instruments used to evaluate the respective library collections. Al though OLIVEIRA (1986)[1] and also PORTA & LANCASTER (1988)[2] compared the results when different CCT measurement instruments were used to evaluate portions of the University of illinois' library collection, only NISONGER (1980, 1983) attempted to test the reliability of several of these instruments.

5 - METHODOLOGY

The two different measurement instruments (MI) to be tested are labeled MI "A" and MI "B" MI "A" is a list of references based on citations

TABLE I. PARAMETERS OF RESEARCH DISIGNIN IN CITATION CHECKING TECHNIQUE STUDIES

Parameters / Authors	sources (MI)	nº of citations	subject areas	nº of samples	scoring method	comp. libr.	tipy libr.	test reliabil.
Bland	25 mono - graphs	250	Total Library	1	Simple % Count	NO	University	NO
Bolgiano, King	Disser-tations 5 year/3 depts.	?	Total Library	1	Simple % Count	NO	University	NO
Buzzard, New	12 disser-tations	1144	Humanities Soc. Scien., Sciences	1	Simple % Count	NO	University	NO
Chambers, Healey	168 disser-tations	7027 2198	English (112 diss.) Education (56 diss.)	1	Simple % Count	NO	University	NO
Coale	4monographs	1883, 733 333, 531	Latin American History Colonial	1	Simple % Count	YES		NO
Emerson	3 biblio. lists 23 dissertations	756	Enginee-ring	1	Simple % Count	NO	University	NO
Gallagher	1monogra phs	100	Ophthal-mology	1	Simple % Count	NO	University	NO
Jewett	4 monographs	139, 251 38, 204	Law, Commece Chemistry, Ethnol	1	Simple % Count	YES	All Types	NO
Lopez1983	5 monographs (Choice List)	100	Sociology, Social, work, Polit. Sci.	1	L o p e s Thechnique	NO	University	NO

TABLE I. PARAMETERS OF RESEARCH DISIGNIN IN CITATION CHECKING TECHNIQUE STUDIES

Parameters / Authors	sources (MI)	nº of citations	subject areas	nº of samples	scoring method	comp. libr.	type libr.	test reabil.
Nisonger 1980	bibliographical lists	100	Mediev. French Lit., Family Therapy , Amer. Novel, Mod. British History	2	LOPEZ TECHNIQUE	NO	University	YES
Nisonger 1983	Periodical Articles	142, 142 150, 150	Political Sciences	2 for each	Simple % Count	YES	University	YES
Oliveira	Per. Artic. Index Journal	220 158	Library Administration	1 for each	Simple % Count	NO	University	NO
Popovich	31 dissertation s	2805	Business Management	1	Simple % Count	NO	University	NO
Porta, Lancaster	Index Journal Documents Derived Journal Articles	500 819 500	Irrigation	1 for each	Simple % Count	NO	University	NO
Webb	4 bibliographical lists	246, 1396 1910, 181	Mediev. Studies Art Hist., Physics Political Scienc.	1	Simple % Count	NO	?	NO
This Study	1 5 monographs 15 Period. Issues	283,282, 234 297, 133, 138	Reference Services	3 for each	Simple % Count	NO	university	YES

randomly selected from monographs on the subjcet of Reference Services, and MI "B" is a list based on citations randomly selected from articles published in journals of the same subject area. Monographs, journal issues, and individual journal articles used as sources for citations were also chosen at random from library holdings dealing with Reference Services.

Monographs and journal articles were chosen as sources of citations which formed MI "A" and MI "B" respectively, because these materials represent two of the three sources for measurement most used in CCT studies. Three samples of citations were gathered to test MI "A" and three to test MI "B". Each sample was compared against the holdings of the University of Illinois's library system in order to determine the percentage held in its collection. This procedure would determine if the three samples within MI "A" and the three samples within MI "B" yield consistent results. The consistency in the results within a specific MI is an indication of reliability

This study will test the following null hypotheses:

(a) There is no significant difference in the results when different samples of citations derived from monographs (MI "A") are matched against a library's collection, and

(b) there is no significant difference in the results when different samples of citations derived from journal articles (MI "B") are matched against a library's collection.

The subject area chosen to evaluate the collection in this study was "reference services". Several factors led to this choice:

1) it is an easy area to identify; 2) it has well delineated boundaries; 3) the library owns a reasonable number of nonographs on the area, leading to better sampling procedures; 4) the library owns enough periodical titles in the area, which allows the sampling of three different sets of references, each from a different title chosen at random.

Five monographs were randomly selected as sources of citations for each of the samples that would test the reliability of MI "A" Since uniqueness of citations was needed in order to test the reliability of the MI, each sample was formed without replacement, that is, once a monograph had been chosen to form part of a particular sample, it was not returned to the original population of monographs from which other samples would be selected

The original list of monographs from which the samples for MI "A" were drawn was derived from a search on the Full Bibliographic Retrieval System of the University of Illinois' Library using the subjcet headings "reference service" and "reference services" This search yielded 104 monographs, of which 39 were relevant. The titles of the other publications carried either the words "reference" of "service" or even both in their titles but with a connotation other than the subject in question.

Since a set of five documents (5 monographs for MI "A" and 5 issues of the same periodical title for MI "B") was used as sources for a sample of citations, a multistage sampling proportional to size of the population was used to form each sample of citations to be checked against the library's collection. Therefore, in order to establish the size of each sample, it was necessary to first identify the total population size of citations in each set of sources (5 items) after the elimination of citations of unpublished materials, duplicate citations within each source, and unidentifiable references. The total population size of each set was identified by adding the number of usable citations contained in each of the five documents that formed a sample.

The size of each sample, therefore, was determined by identifying the population size and its corresponding sample size in the "Table For Determining Sample Size From a Given Population" published by the National Education Association (KREJCIE & MORGAN 1970). Next, the proportion represented by the sample size in relation to the total population was calculated. This percentage value was used to determine the number of citations that was selectsed, at random, from each document that formed a specific set of sources. If the sample size indicated by the Table was 20% of the total population (number of usable citations in all five documents of a set), 20% of the usable citations from each document of the set were randomly selected. The final sample size was obtained by adding the total number of citations derived from each document of the set.

All of the citations were randomly selected from the bibliographies, footnotes and references of the source material. Sample I of MI "A" consisted of 283 citations taken from the first set of five monographs, chosen at random from the 39. Sample II consisted of 282 citiations taken from the second set of five monographs, and sample III was formed by 234 citations extracted from the third set of five monographs.

The three periodicals for MI "B", **The Reference Librarian, RQ and Medical Reference Services Quaterly**, were randomly selected from the five journals specialized on "Reference Service" available at the University's library system. Five volumes of each title were also chosen at random. Next, one issue from each volume was randomly selected. It was the randomly selected issue from each volume that was used as the source of citatons for samples I, II, and III of MI "B". Thus, sample I consisted of 297 citations selected at random from the bibliographies, footnotes and reference notes of all articles published in volumes **5**, **11**, **12**, **14** and **16** (also selected at random) of **The Reference Librarian**. Sample II consisted of 133 citations derived in the same manner, from v.5/n.3, v.12/n.3, v.15/n.1, v.16/n.3 and v.20/n.1 of RQ. The selection of the 138 citations that formed sample III followed the same

procedure. The issues randomly selected from **Medical Reference Services Quarterly** were: v.**1**/n.1, v.**2**/n.3, v.**4**/n.2, v.**5**/n.4 and v.**6**/n.3.

It was assumed the five volumes from each title are representative of all volumes in that title. To this date, **RQ** has published 27 volumes, **The Reference Librarian** 18, and **Medical Reference Services Quarterly** only 6. It was also assumed that any one issue was representative of all the issues in any one volume, which in every case, consisted of no more than four issues.

The citations in the resulting samples were then searched in the computerized and card catalogs of the University of Illinois' Library. The percentage of matches, in each sample, resulting from the search was then recorded by type of material. Since the results of each MI consisted of three independent groups and also represented repeated measurements from the same population, the three results, first from MI "A" and then from MI "B", were submited to the chi saquare test for equality of proportions. By applying this test, it was possible to determine whether the differences among the results of the samples within each MI was significat or not. In order to verify if the differences were due to chance alone, the resulting chi-square value of each variation was then compared to its corresponding critical alpha value at the .05 and .01 level, thus determining the reliability of each variation for the two CCT measurement instruments.

6 - ANALYSIS OF THE RESULTS

The preceding section has outlined procedures followed in gathering the data for this study. This section presents a description of the data and an analysis of the findings resulting from these procedures. The three samples of citations drawn from the monographs will first be presented, followed by an exposition of the results obtained from the matching of the citations against the library's holdings.

6.1 - Results of Measurement Instrument "A"

Table II presents the total number of usable citations and the corresponding sample size, sources by source, for each sample. The five sources in sample I yielded 1093 usable citations. From these, 283 were selected at random, which is equivalent to 26%. This figure was derived by using the "Table for Determining Sample Size From a Given Population." A very similar sample size was arrived at after selecting the usable citations from the sources in sample II. Out of 1102 citations, 282 (25.5%) were randomly selected. Although the total number of useful citations for sample III was approximately half of the number in samples I and II, the percentage necessary to form the

sample size almost doubled, that is, 42%, which yielded 234 citations to be compared against the collection. Adding the results from the three samples, it is verified that a total of 799 citations was selected from 2751 usable ones and matched against the University of Illinois' library collection in order to test the reliability of MI "A" of the CCT.

Table II. Total Population of citations and sample size for the sources in each sample of mi "A"

Source	Sample I		Sample II		Sample III	
	nº usable citations	Sample 26%	nº usable citations	Sample 25,5%	nº usable citations	Sample 42%
1	332	86	243	62	100	42
2	5	1	254	65	152	64
3	230	60	420	107	33	14
4	421	109	117	30	21	9
5	105	27	68	18	250	105
TOTAL	1093	283	1102	282	556	234

Source = Monographs

Since this study proposes only to verify the reliability of the measurement instruments, the differences in sample sizes will not matter, as long as the sampling technique remains uniform for each sample. Only by chance would the number of usable citations and the corresponding sample size for each evaluation be the same.

The variation between the number of usable citations in each source is interesting to note. Considering all sources in the three samples, there is a wide range. Source number 4 in sample I yielded 421 useful citations while source number 2 in that same sample yielded only five citations. These sources are also responsible for the largest and the smallest contributions to sample I, (with 109 citations from source number 4 and only one citation was from source number 2.

The average number of usable citations in the sources forming sample I is 219, varying from 5 to 421 citations. In sample SII the average is 220, varying from 68 to 420 citations and in sample III it is 111, varying from 21 to 250 citations.

The results of the matching process of the citations in each sample against the library's collction are arranged by the type of publication that each

citation represents. Tables III, IV, and V present these results for samples I, II, and III of MI "A".

As Table 3 demonstrates, the library ojwns 80.2% of the 283 documents in sample I. Of these 227 items owned, 52.4%, that is, more than half are journals, 37% are monographs and 10.6% are conference proceedings, dissertations, reports, reference works, and etc. The library holds a very high percentage of the journals cited in this sample - 90.8%. It also owns 85.7% of the monographs and 50% of the dissertations. However, it does not own 53.9% of the conference proceedings and 56.5% of the "other" materials in the sample.

Table III. Results of the matching process sample I of mi "A"

MATCH / ITEM	YES		NO		TOTAL		% of T
	nº	%	nº	%	nº	%	YES
JOURNALS	119	90.8	12	8.2	131	46.2	52.4
MONOGRAPHS	84	85.7	14	14.3	98	34.7	37.0
PROCEEDINGS	6	46.1	7	53.9	13	4.6	2.7
DISSERTATIONS	1	50.0	1	50.0	2	.7	.4
OTHER *	17	43.5	22	56.5	39	13.8	7.5
TOTAI	227	80.2	56	19.8	283	100	100

** - any type of document that does not fall in the other categories.*

- The percetages of columm 2 represents the percetage of items held by the library. The percetages of columm 4 represents the percetage of items not held not the library. The percetage figures of columm 6 are indicative of the percetages that the especific item represents in relation to all other items checked against the collection of the library. The percetage of matches that the specific item represents in relation to all matches.

It is noted that the five monographs that were randomly selected as sources of citations for sample I cited more journal than any other type of document, but also, the library holds a larger percentage of this type of material compared to the other types in the same sample.

As can be seen in Table IV, the five monographs that were chosen as sources for sample II of MI "A", yielded a total of 282 citations, of which the library owns 265, that is, 94%. Of these, 72.1 are journals, 18.9 are monographs, and 9.0% are all other types of materials, including dissertations and conference proceedings.

Table IV. Results os the matching process sample II of mi "A"

MATCH / ITEM	YES		NO		TOTAL		% of T
	nº	%	nº	%	nº	%	YES
JOURNALS	191	96.9	6	3.1	197	69.9	72.1
MONOGRAPHS	50	90.9	5	9.1	55	19.5	18.9
PROCEEDINGS	8	100.0	0	0	8	2.8	3.0
DISSERTATIONS	2	50.0	2	50.0	4	1.4	.7
OTHER	14	77.7	4	22.3	18	6.4	5.3
TOTAL	265	94.0	17	6.0	282	100	100

Although the number of citations in sample I (283) and in sample II (282) are almost the same the library holds a larger percentage of the citations in sample II, 94.0% compared to 80.2%. Just as observed in sample I, the library holds a very large percentage of the journal citations in this sample - 96.9%. It also holds 90.9% of the monographs and holds 50% of the dissertations, as in sample I. Two differences are observed here. While the library holds less than half of the conference proceedings and of the "other" materials in sample I, it holds all of the conference proceedings and of the "other" materials in sample I, it holds all of the conference proceedings and 77.7% (14 out of 18) of the "other" materials forming sample II.

The results of the matching between the citations in sample III of MI "A" and the library holdings are presented in Table V. The five monigraphs that formed sample III of MI "A" yielded a total of 234 citations which were matched against the library's collection. The library holds 80.3%, that is a total of 188 documents. This percentage is almost equal to the percentage held when the citations in sample I were matched against this same collection, (80.2%), although sample I is larger. Of the 188 documents held by the library, 47.3% are journals, 6.7% are monográphs, 12.8% are "other" materials and 3.2 % are conference proceedings and dissertations. As observed in samples I and II, the library holds a larger proportion of the journals cited in this sample than any other material, in this case, 95.7%. This percentage is very close to the one obtained in sample II. The library also owns 86.2% of the monographs, 50% of the conference proceedings and dissertations, and nearly half of the "other" types of materials.

Table V. Results of the matching process sample III of mi "A"

MATCH / ITEM	YES		NO		TOTAL		% of T
	nº	%	nº	%	nº	%	YES
JOURNALS	89	95.7	4	4.3	93	39.8	47.3
MONOGRAPHS	69	86.2	11	13.8	80	34.2	36.7
PROCEEDINGS	5	50.0	5	50.0	10	4.3	2.7
DISSERTATIONS	1	50.0	1	50.0	2	.8	.5
OTHER	24	49.0	25	51.0	49	20.9	12.8
TOTAI	188	80.3	46	19.7	234	100	100

These results are very similar to the findings in sample I. Again it is noted that, as in samples I and II, this sample of five monographs cited more journals than any other type of materials. And once more, the library holds a larger percentage of these as compared to other types of documents cited.

Two interesting tendencies are observed when comparing the results of these three samples. First, the types of documents were cited in the same frequency order in every samples ist journals, 2nd books, 3rd "other" materials, 4th conference proceedings and 5th dissertations. Second, in every case the library ojwned a larger percentage of the journals cited than any other type of materials.

In order to test the first null hypotheses of this study, which is: "there is no significant difference in the results when different samples of citations derived from monographs are matched against a library's collection," it is necessary to compare the total frequency of matches obtained in each sample and test if the differences are significant or not. The chi-square (x^2) test for equality of proportions was applied.

A x^2 equal to 26.71 is obtained when these results are calculated. The degrees of freedom for six cases is 2. The alpha value at a significance level of .05 for 2 degrees of freedom is 5-9 and at a level of .01 is 9,2. The chi-square value, 26.71 is larger then the alpha values, indicating that the differences in the results of the three samples are statistically significant and therefore the null hypothesis is rejected and the alternative hypothesis accepted. The results obtained using different samples of citations drawn from monograhs are not constant, and therefore not replicable. Therefore, MI "A" (i.e. Monographs) might not be a reliable instrument for collection evaluation purposes when applied to the "reference services" area of the University of Illinois'

library, because the differences in the results are statistically significant and not due to chance alone.

6.2 - Results of Measurement Instrument "B"

A second study was developed using similar procedures to test the reliability of another CCT measurement instrument. Whereas for MI "A" randomly selected monographs on Reference Services were used as sources of citations, for MI "B" the citations were derived from articles of periodicals on Reference Services also selected at random from 5 journals specialized in Reference Service held by the University of Illinois' library system.

At first, three journals specialized in "reference services" were randomly selected. They were: **The Reference Librarian, RQ**, and **Medical Reference Services Quarterly**. Each title represents a different sample. Five volume (years) were then randomly selected from each title. The final sample of citations from each group of five volumes was derived from all the articles published in one of the issues in each volume, also selected at random. Table VI presents the total number of usable citations and the corresponding sample size for each one of thefive sources that formed the three samples of citations used in MI "B".

Table VI. Total population of citations and sample size for the sources in each sample of mi "B"

Source	Sample I		Sample II		Sample III	
	nº usable citations	Sample 22.6%	nº usable citations	Sample 66,8%	nº usable citations	Sample 64.2%
1	224	51	84	55	17	11
2	222	50	46	30	101	64
3	166	38	37	24	30	19
4	328	74	24	19	30	19
5	374	84	8	5	37	25
TOTAL	1314	297	199	133	215	138

The five sources in sample I (5 issues of **The Reference Librarian**) of this MI yielded 1314 usable citations. The corresponding sample size indicated in the "Table For Determining Sample Size From a Given Population" is 297, which represents 22.6%. A very different figure was obtained after determining the usable citations in the sources that formed sample II. Out of

199 usable citations, 133 (66.8%) were randomly selected thus forming sample II of references to be matched against the library's collection. A similar sample size was arrived at for sample III. Out of 215 citations, 138 (64.2%) were randomly selected.

By adding the total number of citations in all three samples, a total of 568 citations were selected from 1728 usable ones. The average number of usable citations cited by the five sources that formed sample I is 262.8 (high of 374 and low of 166): In sample II this average is much lower, only 39.8 (high of 84 and low of 8), and in sample III the average is 43 citations (high of 101 and low of 17). The discrepancy in the number of citations contained in the sources in sample I when compared with the sources of the other two samples is due to the fact that the journal used in sample I, **The Reference Librarian**, only publishes entire volume without dividing them by issues as do the other two journals. Therefore, it is natural that it would yield a larger number of articles and consequently, more citations.

This phenomenon did not affect the results of this study because, as can be observed in Tables VII, VIII, and IX, the difference in the results obtained from samples II and III, which had a similar number of usable citations and similar samples sizes, is greater (6.8%) than the difference in the results derived from samples I and II (1.6%) which were formed by completely different numbers of usable citations and also sample sizes. The average sample size of citations derived from the sources in sample I es 59.4 citations, in sample II is 26.6 and in sample III is 27.6 citations.

The results of the matching process between the citations from the **Reference Librarian** that constituted sample I and the University of Illinois' library collection are presented in Table VII. This Table, as all the previous ones demonstrating the results for MI "A", is arranged by the types of documents represented in the samples.

Table VII. Results of the matching process sample I of mi "B"

MATCH / ITEM	YES		NO		TOTAL		% of T
	nº	%	nº	%	nº	%	YES
JOURNALS	165	94.8	9	5.2	174	58.6	62.5
MONOGRAPHS	83	83.8	16	16.2	99	33.3	31.4
PROCEEDINGS	4	66.6	2	33.4	6	2.0	1.5
DISSERTATIONS	2	50.0	2	50.0	4	1.4	.8
OTHER	10	71.4	4	28.6	14	4.7	3.8
TOTAL	264	88.9	33	11 1	297	100	100

The library owns 88.9% of the 297 documents that formed sample I of this MI. Of the 264 documents owned by the library, 62.5% are journals, 31.4% are monographs and 6.1% comprises conference proceedings, dissertations and "other" types of materials.

The library holds a high percentage of the journals presented in sample I - 94.8%. It owns 83.8% of the monographs, 66.6% of the conference proceedings, 50% of the dissertations and 71.4% of all "other" types of materials that formed sample I. The articles that formed the sources of citations for sample I in MI "B" cited more journals than any other type of document. Again, the library holds a larger percentage of this type of material compared to the others that formed this sample.

As can be seen in Table VIII, the articles from RQ that were chosen as sources for sample II of MI "B", yielded a total of 133 usable citations, of which the library owns 116, that is, 87.2%. Of these, 53.4% are journals a37.1% are monographs and 9.5% corresponds to conference proceedings, dissertations and other types of materials, such as reports, reference works etc.

Table VIII. Results of the mataching process sample II of MI "B"

MATCH / ITEM	YES		NO		TOTAL		% of T
	nº	%	nº	%	nº	%	YES
JOURNALS	62	91.2	6	8.8	68	51.1	53.4
MONOGRAPHS	43	81.1	10	18.9	53	39.8	37.1
PROCEEDINGS	3	100.0	0	0	3	2.3	2.6
DISSERTATIONS	2	100.0	0	0	2	1.5	1.7
OTHER	6	85.7	1	14.3	7	5.3	5.2
TOTAL	116	87.2	17	12.8	133	100	100

Table VIII shows that the library holds 91.2% of the total number of journals cited by the sources that formed sample II of MI "B". It also holds 81.1% of the monographs, all of the conference proceedings and dissertations and 85.7% (6 out of 7) of the "other" types of materials.

This sample is characterized by two unique results. It is the only sample of which the library held all of the dissertations in the sample and also the one from which the library matched the largest percentage of the "other" types of materials - 85.7%.

The results of the matching between the citations in sample III and the library holdings are presented in Table IX. The articles published by five different issues of the **Medical Reference Services Quarterly** which formed sample III of MI "B" yielded a total of 138 useful citations. The library holds 111 of those, that is, 80.4% which is the lowest percsentage of ownership when compared to the other samples in this MI. Of the 111 documents owned by the library, 65.8% are journals, 28.8% are monographs and 5.4% are conference proceedings and "other" types of materials. There were no dissertations cited in this sample. The library holds 90.1% of the journals cited, 74,4% of the monographs, 50 % of the conference proceedings and less than half, 40% of the "other" materials.

Table IX. Results of the matching process sample III of MI "B"

MATCH / ITEM	YES		NO		TOTAL		% of T
	nº	%	nº	%	nº	%	YES
JOURNALS	73	90.1	8	9.9	81	58.7	65.8
MONOGRAPHS	42	74.4	11	25.6	43	31.2	28.8
PROCEEDINGS	2	50.0	2	50.0	4	2.9	1.8
DISSERTATIONS	0	0	0	0	0	0	0
OTHER	4	40.0	6	60.0	7.2	7.2	3.6
TOTAL	111	80.4	27	19.6	100	100	100

The library owns a larger percentage of journals cited in all three samples that formed MI "B" than any other type of documents. (except proceedings and dissertations in sample II). The average percentage of ownership per sample in MI "A" is 94.54% and in MI "B" is 92.0% as compared to 87.6% ownership of the monographs cited in the sources that formed the samples in MI "A" and 79.8% in MI "B".

In order to test the second null hypothesis of this study - "There is no significant difference in the results when different samples of citations derived from journal articles are matched against a library's collection", the chi-square test for equality of proportions was again applied. Sample I yielded 264 matches and 33 misses; sample II yielded 116 matches and 17 misses; and sample III yielded 111 matches and 27 misses.

Inserting these results into the x^2 formula, a value of 5.6 is obtained. Since the alpha values of 5.9 at .05 and of 9.2 at .01 level for 2 degrees of freedom is larger than the obtained chi-squared value of 5.6, the differences of the results in each sample are not statistically significant and therefore, the

second null hypothesis is accepted. The results obtained from different samples of citations cited by journal articles are fairly constant and replicable. Therefore, the MI "B" of the CCT tested in this study can be said to be reliable when applied to evaluate the "reference services" collection of the University of Illinois' library.

7 - CONCLUSION

The findings of this study demonstrated a tendency for certain types of material to be cited more often in a specific type of CCT MI. In MI "A" and in MI "B", all of the samples cited more journals then any other type of materials. Journals represented on average 53.0% of documents cited by the sources that formed the samples in MI "A" and 56.1% of the documents cited by sources that formed the samples in MI "B".[3] Due to this pattern of citations, the results of collection evaluations using these two measurement instruments might be biased favorably towards a library which is known to hold comprehensive collections of periodicals, and biased against one which does not concentrate its acquisition efforts in this type of material.

To date, there have been only a few investigations which attempted to compare the results of different CCT measurement instruments. NISONGER (1983) investigated two different measurement instruments using two different samples for each. PORTA & LANCASTER (1988) studied the coverage of the University of Illinois' library on irrigation by applying three different CCT measurement instruments using one sample of citations in each; and OLIVEIRA (1986) evaluated the University of Illinois' library collection on library management using two different measurement instruments with one sample of citations for each. The lack of replication of these studies prevents one from deriving any definitive conclusions about the reliability of the measurement instruments used.

This study narrowed this gap and elucidated some of the questions regarding the reliability of different CCT measurement instruments. By testing two different measurement instruments using three different samples for each and statistically testing the significance of the different results, it was possible to check, the reliability of two of the most common CCT measurement instruments used to assess the strength of library collections.

The results indicated that while citations drawn from monographs might not be a reliable instrument for evaluating library collections, citations drawn from journal articles could be a useful and appropriate instrument for

assessing library collections. This is not to say that the investigation of the problem is completed. It is not possible to make broad generalizations. Further studies should attempt to test these same measurement instruments using a larger number of samples, applied in other types of libraries with different characteristics and also used to evaluate different subject areas.

An aspect that merits further attention is the scsoring method. There are no established criteria for choosing the method by which the results are scored. Almost all of the studies reported in the library and information science literature have simply checked the list of citations against the library's holdings and calculated the percentage of ownership. This scoring method does not take into account good items owned by the library which were not included on the list.

The emphasis should not be only on the overall percentage of matches because there are many variables involved that could influence and alter the final results, depending on their presence or not. For example, number of foreign materials included, comprehensiveness and up-to-dateness of the source materials which yielded the citations to be checked againt the collection, number of years or period covered by the sources, quantity of marginal materials selected, size of the sample used etc. Any one of these variables could alter and affect the results, that is, the percentage of matches or hits between the list of citations and the collection being evaluated. It would be logical to think that a more comprehensive source, or one that includes more foreign materials, or a more up-to-date source, or one that covers a longer period, would yield a greater number of citations less likely to be held by a library than a more specialized or a more domestic source or one that covers a shorter period would.

It is also essential that other mesurement instruments be tested using a uniform procedure lide that developed for this study - citations of documents cited in index journals; citations indexed in secondary sources, such as bibliographies, index journals, literature reviews, as well as citations cited in dissertations. After these studies have been accomplished it might be possible to determine which are the reliable CCT measurement instruments.

Only after obtaining the results of a serie of research studies as described above will librarians be able to answer some of the fundamental questions regarding the CCT, obtaining a better understanding of its process, determining its actual value as a collection evaluation instrument, and defining its real contribution to the library science field. This study is an important step towards the achievement of this goal.

SUMMARY

OLIVEIRA, S.M. *de The reliability of two citation technique measurement instruments.* **Transinformação,** *4 (1,2,3): 55 - 79, 1992.*

The importance of applying reliable measurement instruments when assessing the value of a library collection is discussed in this study. The Test-retest reliability test is used to determine the reliability of two different citation-checking technique measurement instruments - namely, citations derived from periodicals and citations derived from monographs. A chi-square test for equality of proportions demonstrated that periodicals can be reliable instruments while questions are raised regarding the reliability of monographs.

Key words: *reliability, evaluation of collection, measurement instruments.*

REFERENCES

American Library Association. Collection Development Committee (1979). "Guidelines for Collection Development." Perkings, D.L., ed. ALA, Chicago.

Baughman, J.C. (1977). Towards a structural approach to collection development. **College and Research Libraries** 38, 241-248.

Bland, R.N. (1980). College textbooks as a tool for collection evaluation, analysis and retrospective collection development. **Library Acquisitions 4,** 193-197.

Bolgiano, C.E. and King, M.K. (1978). Profiling a periodicals collection. **College and Research Libraries 39,** 99-104.

Bonn, G.S. (1974). Evaluation of the collection. **Library Trends, 22,** 265-304.

Broadus, R.N. (1971). The literature of the social sciences: a survey of citation studies. **International Social Science Journal 23,** 238.

Broadus, R.N. (1977a). Evaluation of academic library collections: a survey of recent literature. **Library Acquisitions 1**, 149-155.

Broadus, R.N. (1977b). The application of citation analysis to library collection building. **Advances in Librarianship 7,** 229-335.

Busha, C.H. and Harter, S.P. (1980) **Researchr Methods in Librarianship:** Technique ahd Interpretation. Academic Press, New York.

Buzzard, M.L. and New, D.I. (1983) Investigation of collection support for doctoral research. **College and Research Libraries, 44,** 469-75.

Chambers, G.R. and Healey, J.S. (1973). Journal citations in master's theses: one measurement of a journal collection. **Journal of the American Society for Information Science, 24,** 397-401.

Coale, R. (1965). Evaluation of a research library collection. **Library Quarterly, 35,** 173-184.

Comer, C. (1981). List checking as a method for evaluation of library collections. **Collection Building, 3,** 26-34.

Emerson, W.L. (1957). Adequacy of engineering resources for doctoral research in a university library. **College and Research Libraries, 18,** 455-460.

Gallagher, K.E. (1981). The application of selected evaluative measures to the library's monographic ophthalmology collection. **Bulletin of the Medical Library Association, 69,** 36-39.

Gove, P.B. (1961). **Webster's Third New International Dictionary of the English Language:** Unabridged. G & C Merriam, Springfield, IL.

Jewett, C.C. (1849). Report of the Assistant Secretary relative to the library. **Annual report of the Board of Regents of the Smithsonian Institute, 3,** 34-47.

Kidder, L. (1981). **Research Methods in Social Relations.** 4th. ed. Holt, Rirnhart & Winston, New York.

Kinnear, T.C. and Taylor, J.R. (1987). **Marketing Research:** an Applied Approach. 3rd. ed. McGraw-Hill, New York.

Krejcie, R.V. and Morgan, D.W. (1970). Determining sample size for research activities. **Educational & Psychological Measurement, 30,** 608.

Lancaster, F.W. (1977). **The Measurement and Evaluation of Library Services.** IRP, Washington D.C. pp. 165-206.

Lopez, M.D. (1969). A guide for beginning bibliographers. **Library Resources & Technical Services, 13,** 462-70.

Lopez, M.D. (1983). The Lopez citation technique of in-depth collection evaluation explicated. **College and Research Libraries, 4,** 251-5.

Martin, P. G. (1952). **Characteristics of the Literature Used by Authors of Books on Political Topics,** Unpublished dissertation, University of Chicago.

McInnis, M.R. (1971). Research collections: an approach to the assessment of quality. **IPLO Quarterly, 13,** 11-22.

Mosher, P.H. (1984). Quality and library collections. **Advances in Librarianship, 13,** 211-238.

Nisonger, T.E. (1980). An in-depth collection evaluation at the University of Manitoba library: a test of the Lopez method. **Library Resources & Technical Services, 24,** 329-337.

Nisonger, T.E. (1983). A test of two citation checking techniques for evaluating political science collections in university libraries. **Library Resources & Technical Services, 27,** 163-174.

Oliveira, S.M. (1986). **Evaluation of the Library Management Collection of the University of Illinois' Library & Information Science Library,** Unpublished paer, University of Illinois at Urbana-Champaign.

Popovich, C.J. (1978). The characteristics of a collection for research in business/management. **College and Research Libraries, 39,** 110-117.

Porta, M.A. and Lancaster, F.W. (1988). Evaluation of a scholarly collection in a specific subject area by bibliographic checking: a comparison of sources. **Libri, 38,** 131-139.

Powell, R.R. (1985). **Basic Research Methods for Librarians.** Ablex, Norwood.

Price, D.S. (1963). **Little Science, Big Science.** Columbia University Press, New York.

Sineath, T.W. (1970). **The Relationship Between Size of Research Library Collection and the Support of Faculty Research Stu-**

dies, Unpublished dissertation, University of Illinois at Urbana-Champaign.

Stewart, J.L. (1970). The literature on politics: a citation analysis. **International Library Review, 2**, 329-353.

Strayer, M.S. (1971). A creative approach to collection evaluation. **IPLO Quarterly,13,** 23-28.

Voos, H. (1981). Collection evaluation. **Collection Building, 3**, 6-11.

Webb, W. (1969). Project COED. A university library collection evaluation and development program. **Library Resources and Technical Services, 13,** 457-462.

NOTES

1. Oliveira used citations from journal articles and citations from documents indexed in an indexing journal to evaluate the UIUC Library and Information Science Library's coverage of library administration. The first sample yielded 88.6% of matches and the second 70%.

2. Porta & Lancaster used citations from IRRICAB, an index journal specialized in irrigation, citations of documents derived from IRRICAB, and citations from journals on irrigation owned by the University of Illinois' libraries. The first sample yielded 50% of matches, the second yielded 86% and the third 78%.

3. In reality, these findings are in contradiction with other findings reported in the literature. NISONGER (1983) found that 362 of the 584 total citations (62%) used in his collection evaluation study referred to monographs, while 222 (38%) were serials. MARTINS (1952) analysis of 3.024 political science citations found 51.3% to be monographs. STEWART's (1970) study revealed that 66% of 1700 citations from Apter & Eckstein's **Comparative Politics** were monographs. BAUGHMAN (1977) calculated that 34.59% of all the citations in the 1974 **American Political Science Review** were serials, where as 65.36% were "non serial". And BAUM et al. (1976) also analyzed the 1974 **American Political Science Review** and report that 59.8% of citations were to monographs, 31.5% to serials and 8.7% to "other" types of documents. It is likely that such percentagaes are subject dependent