

# An analysis of circulation statistics in a university library collection

**Author:**

Condous, Crystal

**Publication Date:**

1985

**DOI:**

<https://doi.org/10.26190/unsworks/7969>

**License:**

<https://creativecommons.org/licenses/by-nc-nd/3.0/au/>

Link to license to see what you are allowed to do with this resource.

Downloaded from <http://hdl.handle.net/1959.4/62513> in <https://unsworks.unsw.edu.au> on 2024-04-30

AN ANALYSIS OF CIRCULATION STATISTICS

IN A UNIVERSITY LIBRARY COLLECTION

Thesis submitted in part fulfilment of the  
Master of Librarianship (Research).

by

CRYSTAL CONDOUS

1985

UNIVERSITY OF N.S.W.

18 MAR 1987

LIBRARY

### ACKNOWLEDGMENT

I would like to thank my supervisor, Professor John Hiller, for his patience, encouragement and guidance. I would also like to thank my family and friends for their perseverance, and a special thank you to my secretary who typed the thesis.

## ABSTRACT

This thesis examines use patterns of circulation data gathered from the University of New South Wales library's external circulation records; it reports on an in-house survey and a limited reserve section room study.

The discussion is related to determining 'core' and to establishing whether 'core' differs between user groups. The thesis surveys the notion of 'core' and addresses the characteristics it should have with regard to stock relegation, weeding and the purchasing of multiple copies.

The thesis proposes an operational definition of 'core'.

The in-house and reserve room surveys were included to provided a basis of comparison with the external circulation data.

## INDEX

	<u>Page</u>
Chapter 1: The Literature.	1
Chapter 2: Data Sources and Collection.	64
Chapter 3: Files and Computer Processing.	114
Chapter 4: Theory Development.	157
Chapter 5: Conclusion.	283
Additional Reading.	308

## CHAPTER 1: THE LITERATURE

	<u>Page</u>
1.1 Introduction.	2
1.2 Overview.	3
1.3 General Theory - Papers - Review.	7
1.4 Obsolescence - Papers - Review.	48
1.5 Major Literature Table.	57
1.6 References.	60

## CHAPTER 2: DATA SOURCES AND COLLECTION

	<u>Page</u>
2.1 Introduction.	66
2.2 Population.	67
2.2.1 Sections of the Dewey Classification to be surveyed.	67
2.2.2 Areas of the collection to be used.	72
2.2.3 Material formats surveyed.	75
2.3 External Circulation.	76
2.3.1 Data collection.	76
2.3.2 Period of data collection.	79
2.3.3 Data analysis.	80
2.4 In-House Survey.	81
2.4.1 Period of data collection.	83
2.4.2 Data collection.	89
2.4.3 Data design.	100
2.4.4 Card format.	100
2.4.5 Sample data.	103

	<u>Page</u>
2.5 Reserve Room survey.	109
2.5.1 Data collection.	111
2.6 Conclusion.	112
2.7 References.	113

## CHAPTER 3: FILES AND COMPUTER PROCESSING

	<u>Page</u>
3.1 Introduction.	115
3.2 Data Files.	117
3.2.1 Circulation Statistical Data Files.	117
3.2.2 Analysis Options.	119
3.3 Aspects of Computer Processing.	125
3.3.1 Processor Requirements.	125
3.3.2 File Development.	128
3.3.3 Ingres Processing.	133
3.3.4 Circulation Relations.	135
3.4 Conclusions.	139
3.5 References.	140
Appendix 3.1 Relational Data Base Structure.	141
Appendix 3.2 Ingres Commands.	148
References to Appendices.	156

## CHAPTER 4: THEORY DEVELOPMENT

	<u>Page</u>
4.1 Introduction.	163
4.2 'Core', its Definition and its Application.	164
4.2.1 Applications for 'Core'.	164
(i) Weeding of Collections.	164
(ii) Relegation.	165
(iii) Loan Periods.	167
4.2.2 'Core' -- A Transactions Activity Distribution Approach.	170
4.3 Short-term Analysis.	174
4.3.1 Distribution.	174
4.3.2 'Total' or Group Analysis	
-- Some Measures.	185
4.3.3 'Total' or Group Analysis	
-- Results.	196
4.3.4 Allowing for different Group Characteristics.	206
4.4 Analysis and Comparison of Results.	210
4.4.1 Results for the Varying Threshold Approach.	210
4.4.2 Results for the Group Characteristic Approach.	212
4.4.3 The Location Problem.	216

	<u>Page</u>
4.5 Long Term Analysis.	226
4.5.1 Distribution.	228
4.5.2 Maximum Transaction Level Analysis.	233
4.6 In-house Survey Results.	239
4.6.1 Survey Limitations.	239
4.6.2 Cluster Analysis.	240
(i) Type 1 -- Several copies used in one day.	243
(ii) Type 2 -- One copy only in use on a given day.	244
(iii) Type 3 -- One copy use over several days (with no date gaps).	244
(iv) Type 4 -- A title with three or more uses over seven days.	245
(v) Type 5 -- Several days use of a title with no more than a 3-day interval between uses.	246
(vi) Type 6 -- One copy use only with at least a 3-day interval between uses.	246

	<u>Page</u>
4.6.3 General Conditions in Clusters.	247
4.6.4 Results of In-house Survey.	247
(i) Conjectures.	248
4.7 Open Reserve Room Results	258
4.8 References.	263

## CHAPTER 5: CONCLUSION

	<u>Page</u>
5.1 Introduction.	284
5.2 Conclusions.	285
5.2.1 Short-term Analysis.	285
5.2.1a Distribution.	285
5.2.1b Total or Group Analysis.	286
5.2.2 Varying Threshold Approach.	287
5.2.3 Group Characteristic Aproach.	288
5.2.4 Long-term Analysis.	291
5.2.5 Reserve and In-house use.	292
5.3 Data Limitations and Future Research.	294
5.3.1 In-house Survey.	301
5.3.2 Reserve Book Room Survey.	304
5.3.3. In Brief.	306

## 1.1 Introduction

This chapter reviews the literature that has significantly contributed to the theory of collection development.

Section 1.2 relates the thesis to the literature.

Sections 1.3 and 1.4 provide two reviews, both of which are on an 'individual paper' basis. These show the role and results of the early theory. Section 1.3 deals with articles associated with the general theory of library use, while 1.4 deals with obsolescence theory. The latter is sufficiently important to warrant separate treatment.

Section 1.5 provides a table which enables the literature to be considered on topic, author, time period, bases.

This chapter treats twenty to thirty references. A list of references for additional reading is included after Chapter 5.

## 1.2 Overview

This chapter reviews the literature which is generally considered to have made a significant contribution to the general theory of collection use in libraries.

The intention is to provide an overview of the library literature which deals with the issues of "library use", including use patterns, obsolescence, stock relegation, collection weeding, and the concept of 'core'.

An analysis of the literature reveals that in monograph and journal use patterns, the following relationships exist:

1. with age, the use of certain types of material diminishes;
2. that circulation use patterns, either current or past, are appropriate indicators of future use patterns;
3. that most use in a library can be attributed to a small core of frequently-used material.

Analyses of circulation use over time have provided the basis for the development of several theories of collection development. These have been prompted by

diminishing budgets and a concern by management with efficiency.

There has been much criticism about the results and their applicability to some libraries. However, the primary purpose of such research has been to establish:

- \* a methodology for removing volumes from collections that have exhausted their usefulness;
- \* a basis for decision-making re the relegation of material to storage, the purchase of additional copies, and the modification of acquisition policies.

Although the literature has been exhaustive in the areas of obsolescence and use patterns, little research has been done on whether the 'core' in a university library is the same for different user groups and whether the concept of 'core' is relevant if the user groups are considered separately.

One aim of the thesis is to examine use patterns of monographs in an undergraduate university collection and to determine whether the 'core' collection is similar for different user groups in a university library and whether the concept of 'core' is relevant if the groups are separated.

The intention is to provide an evaluative tool that will assist in collection review leading to building and maintenance procedures that distinguish between the disciplines and their requirements.

Specifically the thesis will:

1. Determine the shape of use curves for two disciplines in the monograph collection.
2. Determine whether past use histories are predictors of future use in the two disciplines chosen.
3. Determine usage patterns for two different disciplines.
4. Determine whether core collections are similar for different groups of users.

The results of such a study should assist university librarians to answer the following questions:

- \* What holdings would be required to satisfy a substantial proportion of undergraduate needs in a given discipline?

CHAPTER 1: THE LITERATURE

	<u>Page</u>
1.1 Introduction.	2
1.2 Overview.	3
1.3 General Theory - Papers - Review.	7
1.4 Obsolescence - Papers - Review.	48
1.5 Major Literature Table.	57
1.6 References.	60

- \* What are the considerations involved in the purchase of an additional copy of an old title as opposed to the use of more recently published substitutes?
- \* Is the cost of binding (rebinding) of material justified in various disciplines when considered in terms of the use patterns displayed?
- \* What materials can be weeded without serious repercussions to the users?
- \* What material can practically be moved to secondary storage?
- \* Is there justification for separate collections for different user groups?

The thesis also aims to remove a totally subjective approach to the weeding policy of a library, and to minimise the staff time required for the initial assessment of what volumes are likely contenders for weeding. In all collections, some subjective analysis will still be necessary; the purpose is to avoid the handling of all volumes and to direct attention to the titles having the highest probability of being needed.

### 1.3 General Theory - Papers - Review

This section deals with the significant literature associated with the general theory of use.

Philip Morse undertook a study at the MIT in 1968. His purpose was to provide:

"...an effective and efficient procedure for determining how well a library is satisfying the needs of its users."<sub>1</sub>

Morse suggested this could be best achieved by both data gathering and data analysis. His study applied techniques of probability theory to avoid the need for intensive data gathering in future years.

The aim was to provide:

"...a prediction of next year's operation"<sub>2</sub> by "extrapolation of the few items of gathered data; the models then provide the details of the prediction."<sub>3</sub>

Morse suggested that librarians, up until the time of the study, relied heavily on experience and intuition; the availability of quantitative data, using the models he suggested, would assist decisions concerning the services

that should be provided, patterns of attendance, and patterns of book use.

Morse applied probability distributions to library use, and used queuing theory, Poisson distributions and Markov models to look at book use.

These models were then applied in the Science Library at MIT showing both their application and relevance to library collection analysis.

However, Morse did suggest that use of:

"...mathematical models, corresponding to past operations, are of real value only when they can be extended into the future."<sub>4</sub>

His analysis of book use was intended to discover patterns for the present year in order to predict future patterns.

In his analysis of book circulation, Morse found that:

"Thus, the conclusion that our year to year analysis led us, that the circulation behavior of a book depends (on average) on its previous-year's circulation and not explicitly on its age or still earlier circulation enables us to combine the data and improve the accuracy of the results."<sub>5</sub>

Buckland undertook two studies; one in 1970 and, the other in 1975. The first of these was a broad study at the University of Lancaster with the view of obtaining objective information on ways of providing a better library service.

"The study was designed to provide computer-compatible mathematical models based on the 'Bradford Law of Scattering'." <sub>6</sub>

One of the key elements of the research was to determine how long material should be retained. Using a mathematical model for evidence, the study concluded:

"Effective is the policy of retaining heavily-used titles for a longer period than the less heavily used ones." <sub>7</sub>

The study further concluded that it:

"...would be cheaper to satisfy, by inter-library loan, such requests as still occur rather than continue to incur storage costs." <sub>8</sub>

In his analysis of core collections, Buckland reported that:

"20 per cent of the Lancaster stock generates 80 per cent of borrowing." <sub>9</sub>

Buckland also recognised the weakness of subjective weeding and supported the notion that:

"Records of past use are the simplest and best available predictors of future use."<sup>10</sup>

Buckland's second study (1975) analysed features of stock control. He stressed that the study aimed to relate acquisition, duplication, binding and weeding to the needs and behaviour of the library users. The aim was to manage titles that are in heavy demand, as opposed to other use studies which focused on least-used titles and secondary storage. In looking at book availability, apart from the number of copies held, he suggested other critical factors determining availability were:

"- the frequency with which the book is sought (its popularity), and

- the length of time it is off the shelf when used."<sup>11</sup>

He concluded that the following relationship existed between factors associated with loan policies:

"- For any given loan period, the chances of a copy being on the shelves when sought varies inversely with the popularity of the book. The greater the popularity, the lower the immediate availability; the

less the popularity, the higher the immediate availability.

" - For any given popularity, the length of the loan period and the immediate availability are inversely related. The longer the loan period, the lower the immediate availability; the shorter the loan period, the higher the immediate availability.

" - For any given level of immediate availability, the popularity and the length of the loan period are necessarily also related. The greater the popularity, the shorter the loan period has to be; the less the popularity, the longer the loan can be.

" - Duplication. Increasing the number of copies available, like shortening the length of the loan periods, increases immediate availability..."<sup>12</sup>

To make material available promptly Buckland suggested:

"...the cardinal rule of library stock control is that both the loan period and duplication policy should be related to the level of demand for the title and to each other."<sup>13</sup>

Buckland extended the study to look at loan and duplication policies for open shelf collections.

Open shelf (access) collections have three major problems:

1. Lack of a single loan period.
2. Difficult data collection.
3. The importance of reservations.

Buckland used Monte Carlo simulation in his analysis. Instead of more conventional solutions, e.g. secondary storage, Buckland introduced variable loan periods for the University of Lancaster Library. Items having more than three loans per year were categorised as 'popular' and given a shorter loan period. To remain in the popular category the items had to be issued four or more times in one year, otherwise they were demoted.

Buckland's variable loan and duplication policy, on examination, revealed strong support for the theory that:

"...the demand for library services is elastic... in that (improvements) in the provision of services will increase demand."<sup>14</sup>

Improvement in book availability at Lancaster was followed by an increase in book usage.

Buckland indicated that the use of documents follows certain patterns. The most relevant patterns for managing library stock, according to Buckland are:

"(a) The distribution of demand over titles in a pattern usually known as Bradford's Law of scattering;

(b) The distribution of demand over time in a well-established pattern of obsolescence." <sup>15</sup>

Using these for analysis is not without difficulties, especially in data collection and behavioural response of users.

Lister, in 1967, focused his study on the costs involved in compact storage. He accepted findings that the best two predictors of future usage are age and book usage rates. He studied the economic impact that resulted when these two criteria were used to select books for compact storage. Lister looked at three branch libraries at Purdue University using a 20 per cent sample of the titles from the shelf-list and tabulating use and age data by computer.

Lister believed that:

"...selection of an item for storage should be based entirely upon its current (or immediate past) rate of usage...age criterion...was found to be far inferior to the usage rate criterion for scientific monographs." <sup>16</sup>

Lister found that there was evidence which illustrated variability in the:

"...current usage rates of books of a common age, demonstrating the infeasibility of a single age-related obsolescence function..."<sup>17</sup>

His reasoning for declaring that the age criterion was deficient was:

"The age model is based upon the average behavior of obsolescence with age...The variability about the average rate of usage for each age group is completely ignored. It is because this dispersion about the mean is often very significant that causes some authors to protest a storage decision rule predicated upon the ages of books and to support the usage rate criterion." (sic)<sup>18</sup>

Cost was the major focus of the study and formulae were developed to measure total cost for decision-making. The weeding criterion was a secondary element in the study.

Trueswell's thesis (1964) approached the weeding problem from the point of view of the applicability of data processing and computer techniques. The emphasis was on behavioural patterns and user requirements in a university library.

The other part of the study, using current circulation samples, determined:

"...circulation rates, charge data distributions and book age distributions"<sup>19</sup>

The purpose was to establish a core collection. Based on the previous literature that only a small portion of a library's holdings are in circulation very frequently, Trueswell suggested that this subset of volumes was core and defined it as:

"...core collection as a percentage of the collection that should satisfy a given level of the user circulation requirements."<sup>20</sup>

He undertook in 1964 an analysis designed to evaluate the need to prepare punched book cards for all volumes in the library. The data used was circulation figures from the Deering Library and the Technological Institute Library of Northwestern University. His analysis considered two factors: the length of time the book had been in the library and the last previous circulation date. In relation to the age factor (based on accession date), he made broad assumptions about future circulation patterns and assumed there would be no change. However, he did indicate that additional data would be required.

The last circulation date factor was seen as a more productive approach. From his sample at the Technological Institute Library, he concluded that:

"...74 per cent of the sample had been through the circulation control function of the library one or more times during the preceding five-month period."<sup>21</sup>

He extended the argument to consider the value of mechanised procedures for circulation using the 'frequency of book use' concept. Given his Technological Institute sample, he indicated that, had a circulation system been put into operation 18 months previously,

"...one could expect...only 9 per cent of the books currently borrowed to be lacking book cards."<sup>22</sup>

He went on to assume that the frequency distribution of the circulation sample was representative of future circulation. In so doing, he put forward the following analysis about loan patterns:

"Thus the statistic of the previous charge date apparently gives a good indication of the time and workload requirements for the implementation of a punched book and circulation control system..."<sup>23</sup>

He concluded that only a small conversion workload of the daily circulation would be necessary. His Technological Institute Library example reveals a three per cent current conversion after the circulation system had been in operation for 36 months.

One of his more important findings was that:

"It indicates a very small portion of the library's holdings is accounting for a large fraction of the daily circulation."<sup>24</sup>

In addition, he saw the last circulation date as having:

"...potential in the preparation of a decision rule for a quantitative method of thinning stacks."<sup>25</sup>

Although not stated, it may be assumed that since he was discussing circulation, he referred to volume thinning not titles. He concluded that the last circulation date may be used to:

"...define quantitatively a core collection of well under 50 per cent of the present holdings that will satisfy over 99 per cent of user circulation requirements."<sup>26</sup>

He did not, at this stage, define 'core' and suggested that such a definition was in progress.

He saw other implications resulting, e.g. multiple copy needs, non-circulating volumes, and size of library's holdings being quantitatively measured.

He concluded that the future could see restricted collection growth with no loss to user needs.

Trueswell's extension of this work (1965) looked at a methodology for weeding collections based on the last circulation date and the focus was on monographs.

He suggested that:

"...the criteria for stack thinning should be designed to help the library satisfy the requirements of the users of the library."<sup>27</sup>

The relationship that he established used the percentage of the circulation sample circulated within a certain time frame. If, in the sample of 36 months circulation, only three per cent had not been charged out, then if the books not circulated after the given time period were discarded, three per cent of user needs would not be fulfilled.

He developed a decision rule for the "cut-off date" for weeding collections which suggests:

"Remove all books that have not been circulated during the previous eight-year period."<sup>28</sup>

In this analysis Trueswell did not consider either in-house use or browsing and further assumed that the current circulation frequency distribution patterns would be valid for the life of the library's circulation.

In his next work (1966) Trueswell re-affirmed the findings of his previous research and suggested that:

"...last circulation date may be an ideal statistic to define and measure circulation requirements and patterns."<sup>29</sup>

He assumed that the circulation pattern, as measured at the present, is typical of the circulation patterns of 5, 10 or 12 years in the past.

The most significant of his findings was that:

"...it may be possible to predict the subset of the library's holdings (core collection) that will satisfy a given percentage of user circulation requirements."<sup>30</sup>

Trueswell did not take the argument further. Debate has developed around the core/circulation concept. Having determined core and nominated the percentage of transactions that can be defined as core, e.g. 80-90 per cent, the Trueswell analysis was based on the concept of a full collection, hence the specified titles would account for the percentage of core as indicated. However, if non-core material is removed, the remaining titles would not necessarily handle as many transactions as would be the case with the full collection. Ultimately, each core redetermination over time would reduce the use of the collection though it may increase its efficiency.

In a study undertaken at the AF Cambridge Research Library in 1968, Trueswell looked at the percentage of users accounting for corresponding circulation. The results indicated that a high proportion of the circulation could be attributed to a low proportion of active borrowers.

The collection was in the physical and environmental sciences and contained 125,000 volumes. The user population was approximately 1,000 professional people (scientists, engineers, laboratory staff). Of these, 835 borrowed at least one book. Although the results indicated a pattern that would be expected, Trueswell's study provided a quantitative picture. The analysis was, however, undertaken in a particular library. Trueswell

suggested that results from other libraries may not be similar and would require further analysis.

In 1969 Trueswell studied two more libraries: Holyoke College Library and Goodwell Library, University of Massachusetts, in addition to the Deering Library. Although all three showed similar stack/circulation curves, the two new libraries were far more comparable. He did, however, indicate that the sample sizes were only a small proportion of the libraries' holdings.

Trueswell indicated that by using stack/circulation data:

"...it is now possible to define core collections at any desired percentage of circulation satisfaction."<sup>31</sup>

He provided a methodology for identifying core collections where there were no adequate circulation records. He concluded that more work was required in:

"...evaluating the use of the last circulation date cumulative distribution as a parameter for defining and predicting library circulation patterns."<sup>32</sup>

He re-affirmed its use in identifying multiple copies, automated circulation system introduction, high/low use material and weeding of volumes from collections. In a

more detailed analysis (1968) Trueswell re-affirmed the use of the:

"...last circulation date as a statistic to help describe library user circulation requirements."<sup>33</sup>

He recommended use of circulation data for decisions on purchasing multiple copies and for identifying material for weeding.

The approach of cumulative distribution functions of last circulation dates takes out the subjective elements in weeding collections, and

"...allows a quantitative approach to an otherwise subjectively treated phenomena."<sup>34</sup>

He restated his earlier advice:

"...there is a group of books that could be called a core collection that circulates quite frequently and... only a very small percentage of circulation represents books that have not circulated within a relatively short time period."<sup>35</sup>

None of the research undertaken by Trueswell looked at the concept of 'core' in relation to differing user characteristics, and he concentrated on what in this thesis is termed 'total core'.

Jain (1967) in his detailed study on book use was concerned with the inadequacies of the data collection techniques for book usage studies. He considered two methods, the "collection sample" and the "check-out sample". The former provided information on usage over a longer period of time with results valid for the entire library collection. Jain concluded that:

"It is difficult to design and operate a good sampling plan for data collection."<sup>36</sup>

This method has problems of missing data and the interpretation of circulation usage as circulation rules change in libraries over the years. The main problems of the checkout sample method arise from the fact that it relies on a sample of books charged out over a period of time. Consequently it may be difficult to extend the findings to the total collection and it is not possible to obtain information on the rate of usage of the same group of books over a long period of time.

To minimise the disadvantages in both methods and to maximise the advantages of each method, Jain developed the concept of "relative use".

He defines this as:

$R_i$  = percentage of relative use of the  $i$ th group

$$= \frac{H_i + I_i}{S_i} \times (100)$$

where:

$S_i$  = number of titles in sample S which belongs to the  $i$ th group

$H_i$  = number of titles in sample H which belongs to the  $i$ th group

$I_i$  = number of titles in sample I which belongs to the  $i$ th group

Each is an independent sample of monograph titles obtained from:

- (i) Total collection (S)
- (ii) Home used collection (H)
- (iii) In-library used material (I)

The samples are then further divided into a number of groups on the basis of the following characteristics:

- "(i) Language
- (ii) Country of publication
- (iii) Year of publication
- (iv) Year of accession"<sub>37</sub>

In his analysis using all three methods, the relative use method estimates were much closer to the true values than the other methods. Hence he concluded:

"Sampling methods and the concept of relative use developed in this research study, are more efficient in studying usage of books than the usual method employed."<sub>38</sub>

Jain reviewed the mathematical models employed in book usage studies and indicates that all are deterministic in nature, except Morse's Markov model. (This suggests use of a book follows a Poisson distribution.)

In Jain's analysis he found that "...zero-use class does not follow the same probability law as the remaining classes".<sub>39</sub> To deal with this, Jain developed his Pn model which emphasises the probability of a book not being used.

Jain's other concern was the over-emphasis on usage histories in book usage studies and he commented:

"...that age is a significant variable in studying usage of monographs."<sup>40</sup>

In addition, he observed:

"While usage rates of individual monographs have considerable variation even over a short period of time, the usage rates of various age groups do not show significant differences over time."<sup>41</sup>

Jain's major aim was to examine mathematical models for studying the dependence of circulation rate on age.

Fussler's and Simon's (1969) research aim was to find a characteristic or variable that would predict which books were most likely and least likely to be used.

Data was gathered from the entire collection, as distinct from current circulation (which is obtainable at the loan desk).

They added the concept of "cross-sectional approach", as total data was unavailable, and recorded use in predetermined periods of time. They assumed future usage would continue at the same rate as in the past. This use

ranking was then related to the variables to be studied, e.g. age, use, year since last use. The best function is then considered to be the one which identifies the fewest books in the core which can maintain a predetermined level of use.

The major findings relating to shelf-time period and the age of the volume were:

1. Using years since last use as the only variable gave good results.
2. Age of the book and language characteristics are less satisfactory in predicting future use than in past use.

Past use, where sufficient data are available, was found to be the best single predictor of the future use.

However, it was implied that differences in libraries necessitate individual study for each collection.

McGrath (1975), although agreeing that circulation variability in subject areas is partly dependent on the teaching program, suggested that if the number of students enrolled is the same, and variations still exist in circulation, that the variation could be attributed to the nature of the subjects themselves.

In his study he considered categorisations of hard/soft, pure/applied, and life/non-life to subjects and determined

the relationship between these characteristics and the number of books charged out of an academic library.

The hard/soft distinction is developed out of Kuhn's theory of paradigm development. According to Kuhn a paradigm:

"...stands for the entire constellation of beliefs, values, techniques and so on shared by the members of a given (scientific) community."<sup>42</sup>

According to Biglan:

"...a high paradigm, or well developed discipline, would be 'hard' and a low paradigm discipline would be 'soft'."<sup>43</sup>

Using the premise that a high paradigm discipline communicates through journals and a low paradigm discipline through books, McGrath suggested that it could account for differences between disciplines in book and journal use.

McGrath defined pure/applied as being applied = action; pure = knowing, life as being life systems, e.g. biological sciences, and non-life representing abstract, e.g. mathematical concepts.

McGrath's major findings were:

The hard/soft characteristic was significantly correlated with subject book use when the number of books held by the library and masters enrolments were held constant. However, when strong variables are not controlled, there is no significant correlation between subject book use and hard/soft. If the pure/applied is controlled, McGrath did find a significant correlation, accounting for 11 per cent of the variance.

There was no support for the hypothesis that the more applied the subject, the fewer the number of books borrowed and conversely the purer the subject, the greater the number of books borrowed.

McGrath also could not support life-oriented subjects as having more books borrowed than non-life oriented subjects.

Urquhart and Urquhart's study (1976) at Newcastle-upon-Tyne was looking for procedures for relegating stock that would be applicable to all university libraries. They, like McGrath, found that the difference in circulation patterns in different subject areas were not significant. Hence, there was reason for using a uniform cut-off point for all subjects in the area.

They also found that recently acquired stock is most used in academic libraries.

The most recent use study is that of Allen Kent et al. (1968-1975) known as the University of Pittsburg Study.

The Kent study examined both book and serial use in a research collection, emphasising the humanities and social sciences. The purpose was:

"To develop measures for determining the extent to which library materials (books/monographs and journals) are used, and the full cost of such use."<sup>44</sup>

The study looked at complete acquisition and transaction records for monographs made by the library from 1968 to 1975. It looked at items rather than titles, but suggested that there were fewer than 5 per cent of titles with duplicate copies available. Although the use of unique items will give some 'error', it was not considered to be significant to their findings.

Their results support the hypothesis that a small proportion of the library collection of book titles accounts for the major portion of circulation use, 51.63 per cent of items circulated externally during the period, i.e. 48.37 per cent were not used through external circulation. The other emphasis of book use was to look

at use in terms of newly acquired materials to discover the relationship between use and time. Some of the questions addressed were:

Are they used at all?

If so, what time frame?

How does use change over time?

Can probabilities be assigned for materials that have never been used after a certain time period?

Fussler and Simon, Trueswell and Morse have looked at these questions using only sample data. The Kent group attempted to analyse full population data. Their approach was historical with the intention of developing sampling methodology for future studies.

The study found that a definite aging pattern emerges. As books get older, first time use is reduced in each following year. By the end of 1975, 60 per cent of the 1969 acquisitions had been used at least once. The greatest increases occurred in the first two years. It was found that each year's acquisitions behave much like any other year; the percentage of the new items entering the circulation collection peaks in the first year, then rapidly declines.

This suggested that the distribution is some sort of exponential function that approaches an asymptote. The Kent team developed a model using asymptotic regression to predict the use of acquisitions. They found that:

"The model appears to predict more accurately for the years earlier in the life of a year of acquisitions than it does for later years."<sup>45</sup>

They attribute this to the availability of more data for the earlier years, and suggest that for the later years the collection had not aged sufficiently. They go on to suggest that:

"...the potential circulating collection that will ever be used will settle somewhere between 60% and 70% of acquisitions for any single year."<sup>46</sup>

Their data also revealed that:

"...75-78% of those books and monographs that circulate in-house also have circulated externally."<sup>47</sup>

and it was concluded:

"...in terms of whether or not a book or monograph is ever used, it is sufficient to examine the external patron circulation data."<sup>48</sup>

It was found that 84 per cent of reserve book room transactions (i.e. charges to reserve and not loans in reserve) also circulate externally, and further that the items were high-use items in their own right. Reserve book room materials average 7.2 external circulations per item to their reserve book room transaction.

In the study, frequency of use analysis supported the hypothesis that a small portion of the book collection accounts for total use and further that:

"...a one year period shows a log-normal distribution and is, in fact, an almost straight line when graphed on a semi-log scale."<sup>49</sup>

It also revealed that:

"...the number of items that circulated twice, is approximately half the number that circulated once; the number that circulated three times is approximately half the number that circulated twice and so on for a fairly regular progression."<sup>50</sup>

It goes on to say:

"It appears that circulating books form a Bradford set and that the distribution of circulation over a population of items is a Bradford-Zipf distribution."<sup>51</sup>

The Kent group advocated objective techniques for weeding. Like Trueswell they included in 'core' any item that has circulated externally one or more times. Once the item enters 'core' it remains there. The potential circulating collection is defined as containing core and items never circulated. The model was seen to be useful in moving items to secondary storage as well as weeding.

Slote, whose work spans the period 1969-1980, tried to establish variables that would be useful and provide meaningful criteria for weeding. In his original study (1969) he examined the use of 'shelf-time period', i.e. "...the time a book remains on the shelf between successive uses; and the age of a book as indicated by the most recent date printed on the title page...'imprint age'." <sup>52</sup>

The study also aimed to establish a better criterion, i.e. "...which would yield a smaller core collection that would satisfy a given level of predicted future use." <sup>53</sup>

The third aim was to indicate whether the "...pattern of use of volumes currently in circulation...was as valid a predictor of future use as are historical reconstructions of usage over much longer periods of time." <sup>54</sup>

His findings were:

"1. Past use patterns...are highly predictive of future use.

"2. ...'imprint age' is a weaker predictor of future use than 'shelf-time period'.

"3. The 'shelf-time period' is a predictor of a 'better' core collection...because it describes a smaller core collection for the same level of predicted future use."<sup>55</sup>

("Better" is defined in terms of a 'smaller' core. Further, he suggested that shelf-time period patterns also predict future use patterns.)

"4. Each library tends to have its own unique patterns of circulation and each needs to be studied individually.

(Slote suggests that no weeding criterion can be applied from one library to another, although he gives no explanation why.)

"5. That the 'current circulation' method...[is] as valid for predicting future use as is the method of historical reconstruction."<sup>56</sup>

His application of the criteria in the Harrison Public Library revealed the following:<sup>57</sup>

1. Using shelf-time period as a criterion, circulation actually increased in the area weeded, and did not reduce circulation as expected. The rest of the collection did not show similar increases to the weeded area.
2. Shelf time varied between classes of books.
3. The library's holdings were not in proportion to the uses being made in those classes.

(Slote suggested that there could be improved optimisation of book investment based on use.)

4. Monthly variations called for caution. Slote suggested a sample of at least 400 cases and indicated that final weeding should not be undertaken too rapidly.
5. Non-core volumes retained should have been weeded. This applied to those volumes maintained as a result of professional subjective criteria.

In his third study, he compared [the] "'current circulation method' of creating shelf-time periods, from a

sample of the most recent week's usage, with the 'historical reconstruction method', which creates shelf-time periods from a sample of 10 years' usage".<sup>58</sup> The study revealed that no significant differences could be detected and confirmed that usage patterns do not change over extended periods of time at levels that might be used for weeding.

Voigt (1979) criticised Kent on the basis that circulation is not an adequate indicator of research use in a research library. He suggested that research use can only be studied by concentrating on that research use, not on the statistics that result from intensive use of a few volumes by undergraduates.

Voigt suggested that research patterns are not the same as undergraduate patterns. He stated that there are basic assumptions which must be considered in the study of research use and which must be disproved before being discarded.

The assumptions are:<sup>59</sup>

- \* That in-house use of research material to circulation may be as high as 50 to 1.
- \* Books circulated to faculty have a low correlation with material they use in-house for research.

- \* Faculty and researchers do not leave books on return shelves or tables.
- \* Research material circulates infrequently.

Voigt criticised the Kent team for discarding these assumptions without proof and for basing their work on the assumption that research use follows the same patterns of use as that of undergraduates, which he suggested was simplistic.

Voigt was also critical of the journal analysis, suggesting it did not give reasons for why a periodical is used, nor did he feel that the Kent team had allowed for multiple in-house use.

Voigt agreed that the study reinforced what had been found previously, i.e. that:

"Academic library circulation is concentrated in a small number of books withdrawn primarily by undergraduates and beginning graduate students." 60

However, Voigt went on to state that:

"(Kent)...can contribute little to our knowledge of collection development or maintenance of what the

limits of resource sharing are for universities with heavy commitments to research." <sup>61</sup>

Massman (1979) suggested the Kent study would make library administrators more aware of what they were doing, and would force them to re-examine their theoretical base for collection development.

He indicated that reliance on circulation studies has pitfalls. For example, it is difficult to predict what will be of value ten years hence because of changing issues. The missions of the institutions also vary producing differences in the libraries which serve those educational missions. He pointed out that access through inter-library loan is now much easier than ten years ago, so libraries have no need to collect everything. However, there must be a base; inter-library loan is not a substitute for quality local collections.

Massman said that by definition research materials are low use materials and inter-library loan access is no substitute for the scholar's access to books and journals.

Although the principle of supplying current interest materials in a small public library may work, the principle "...[applied to] academic libraries [will] severely limit the possibility of real scholarship and research." <sup>62</sup>

Massman concluded that:

"...circulation statistics are a useful management tool, but their application still requires all the knowledge, judgement and wisdom we can muster."<sup>63</sup>

Trueswell (1979) in reply to Schad's criticism of Kent indicated that there were three issues:

1. "The results of the research study,
2. "the effects of these results on library budgets,
3. "the appropriate size of library budgets."<sup>64</sup>

Trueswell stated that, even if there is criticism re Kent's data collection, the numbers are secondary.

He noted that it is clear that the research, Kent's, his, and others' reveals:

- "(1) There is a fairly large number of books that are never used, and
- (2) most circulation use comes from a small portion of the holdings (80/20 rule, etc.)".<sup>65</sup>

Trueswell felt the central issue revolved around the objectives of a library and how one should react to them.

If it is a research library, then it should have everything; if it is for undergraduates, then they can be served with fewer volumes. The problem is that most libraries have a mission that is between the two extremes.

Another problem with the research is that it indicates user behaviour after the event and does not indicate what ought to be purchased. He argued that research libraries have two types of use: one of instruction; the other research.

Trueswell suggested that research to determine the magnitude of each of these two kinds of use in a library, perhaps by subject, might shed more light on the financial support needed for collection building of each type of use. He states that:

"...infrequently used books must be made available to all users through one medium or another...what medium should be used is the question."<sup>66</sup>

Trueswell commented:

"What needs to be done is to use the results of these studies in conjunction with current technology and systems analyses, to meet most, if not all, of the needs of library users."<sup>67</sup>

Schad's criticism of Kent (1979) centred around the conclusions of the study. First, he criticised the data used. Schad suggested it was incomplete, e.g. 1969 acquisition figures were inaccurate and did not include all acquisitions as losses were not included; the loans manually charged were not included in the study, which was as high as 5 per cent and, finally; the unaccounted misreads in the automated system had also been excluded.

Schad criticised the assumption stated by the Kent group that external use is an adequate measure of total use. He indicated that Kent's study showed that a large number of books not circulated were used internally. The converse statement is made in the report. The study does not deal with returns to the shelf by the user. Schad suggested that accurate figures re in-house use cannot be obtained unless this in-house use is studied in more detail.

Other criticism made by Schad was more subjective and centred around the role of the research library.

He mentioned two types of use in academic libraries. One is instruction, the other is research. He indicated that the two cannot be measured together. Based on this premise he considered the use of a single instrument to evaluate library use inappropriate.

He defined instructional use as providing intensive use dependent on a small body of material, often requiring multiple copies. Research use is less intensive and relies on a broader range of materials. The other characteristic that Schad said had been ignored was significant changes in research interests and the changing composition of faculties, both of which will often result in significant shifts in the use of titles.

Finally, Schad suggested that:

"...collections for research are built knowing that many volumes may not be used heavily either immediately or at any time...expenditures for research materials must be considered as an investment in that funds are spent in the hope of realizing a return in the future and the knowledge to be obtained cannot be specified." 68

Kent's rebuttal (1979) was directed at Schad. He questioned Schad's interpretation of his conclusions which Kent states are the results of the data analysis. The conclusions Kent suggests relate to opportunities for weeding and low cost storage of material which is not high use; resource sharing and tracking use over time as an aid in collections' development.

Kent indicated that the missing data re 1969 acquisitions arose because the study took a conservative approach and

based the data on calendar, not fiscal, year acquisitions. (The annual report figures quoted by Schad relate to the fiscal year.)

He also pointed out that the manual charges were for pre-1969 acquisitions and were not part of the study. In terms of misreads the error was calculated at 2-3 per cent, the use distribution of this percentage would not be likely to differ significantly to the other 97-98 per cent. He further added that this percentage (97-98 per cent) must be indicative of total use.

The in-house use criticisms levelled by Schad are also rebutted by Kent. The reshelfed book argument, Kent claimed, is a matter of opinion. The extrapolation to later years, according to Kent, is possible provided there is continuous analysis.

The relationship of in-house use to circulation, Kent suggested, was a limited statement indicating that external circulation data can be utilised with a high level of confidence to measure total book use in terms of books used at least once.

McClellan (1978) introduced measures for stock control purposes. The system he discussed was clerical, but computer-adaptable. McClellan suggested that the model was both theoretical and pragmatic.

His purpose was to propose a methodology that could be applied which would both give priorities and maintain stock qualitatively and quantitatively at the best levels the resources would permit. His methodology attempted to ensure that less popular categories of stock achieve a viable level of provision for smaller readership groups without penalising the more popular levels.

McClellan does this by the maintenance of desirable stock levels and introduces square root weighting. Within the total bookstock "...[it] introduces a bias which takes account of the need for the smaller interest groups to have a larger selection to choose from than their actual size would warrant on a strictly proportional basis".<sup>69</sup>

In applying the square root method, McClellan goes further and develops a formula for desired shelf stock, which is:

$$T_s = \left( \frac{S\sqrt{(L_1)}}{R} \right)$$

where:

$T_s$  = target shelf stock for an Interest Category

$S$  = total circulation shelf capacity of library

$L_1$  = highest number of books on loan as shown by the issue 'counts' of the previous year

$R$  = total all square roots ( $\sqrt{\phantom{x}}$ ) for all Interest categories ( $L_1$  to  $L_n$ )<sup>70</sup>

In addition, McClellan considered in his analysis depreciation rates (physical decay) and obsolescence and developed a formula for an annual replacement target.

McClellan tried to achieve the following with his method:

"Essentially, it is concerned with sensing the condition of the stock throughout all its categories to ensure that they are maintained at a continuously uniform level of effectiveness consistent with their varying current and potential readerships."71

He further stated that the method must be used in conjunction with the aims and objectives of the total library system and, where possible, in its own situation.

"...to see itself (the library) as one segment of an entire retrieval system for the whole field of literature and aiming within the segment to offer the highest levels of accessibility possible thereto for as many readers as possible."72

Finally, McClellan stated that:

"...(the method) is designed not only to minimize the occurrence of unsatisfied demand by setting up alarm

signals about the condition of the stock, it is also sensitive to any demand for the greater representation of an 'interest' if not of specific titles."73

McLellan stated that it was not a method of book selection but was concerned with the quantitative aspect of book provision.

"It takes into account, the number of readers and the distribution of their interest...the shelving capacity of the library, the freshness of the stock, the accumulated actions of the book selector to date and the costs of replacement of the stock."74

Although in some of the use studies examined in this chapter use-obsolescence in time-related aspects of collection use are discussed, there is a body of literature that deals specifically with the obsolescence of library materials. It was felt that this area relating to collection use should be discussed separately.

#### 1.4 Obsolescence - Papers - Review

This section deals with papers that have contributed to the theory of obsolescence.

The first significant paper on obsolescence of library material was due to Maurice Line. This introduced the following measures:

- "1. Apparent obsolescence factor.
2. Corrected half-life.
3. Corrected obsolescence factor.
4. Half life.
5. Item half life.
6. Median citation age."<sup>75</sup>

These measures were applied to use curves of serial literature to ascertain obsolescence rates.

Apart from these measures, Line also suggested that a direct relationship existed between the growth rate of the literature and obsolescence rates, adding a further dimension to the obsolescence theory.

Brookes (1970) pointed out that the concept of a relationship between the growth of literature and obsolescence rates required techniques that were so precise that they were probably unattainable.

His concern was with Line's sampling techniques and his measurement of the growth of literature.

Brookes also suggested that Line's measures had omitted the inclusion of what he considered to be an important factor. He called this the utility factor, i.e. the need to specify the group of users, because of differing needs of particular user groups in varying situations.

He further suggested that to obtain a rate of obsolescence for a particular library it is not sufficient to 'correct' Line's general measure (as suggested by Line). He states that what is required is the sampling of the actual usage of the literature within the library. If this is correct, then it must be assumed that a general theory may have limited applicability.

Brookes also draws attention to the importance of the type of study undertaken. He suggested that synchronous studies, as opposed to diachronous studies, would reveal quite different results.

A further extension of the Line/Brookes theory has been made by Buckland (1972). A small study undertaken by Buckland, suggested a relationship between Bradford's Law and obsolescence rates. If such a relationship exists, core collection material would become obsolete at a slower rate than non-core collections. If this is so, then

smaller collections would not seriously affect the quality of service. Another aspect that arises out of this concept is, if either the obsolescence rate or scattering is known, can the other, which is not known, be determined? If this is proven to be possible, then a further dimension will be added to the obsolescence theory.

The measures introduced by Line and Brookes have been used in a number of studies, usually in an attempt to look at a specific problem the library was facing at the time.

Chen (1972) based her study on usage of journals and the "half-life" concept with the intention of setting up a core collection of physics journals which would satisfy 90 per cent of the users' needs.

Rao (1973 and 1974) undertook two studies, one in the social sciences and the other in engineering. The aim was to determine whether the obsolescence patterns were the same in the two areas chosen. He used all the measures as put forward by Line to determine these patterns.

Kohut's study (1974), apart from using the basic Line measurements, extended his study further by incorporating Line's concept of the growth of literature into his methodology.

A study on agronomy journals by Lawani introduced a new concept which has yet to be proven. He suggested that, if the subject area chosen is extremely narrow, then the Line/Brookes measures are not fine enough for accurate measurement.

There are several important aspects of these studies:

1. All were undertaken for serial collections.
2. Results show heavy usage in the first two or three years and slow decay after this point in the hard sciences. The social sciences, on the other hand, have a much slower decline. This indicates that the hard sciences have a comparatively short "half-life"; the social sciences a long "half-life".
3. All the studies reveal exponential decay and this has been disputed only once. A study by Line and Carter (1974) revealed linear decay. Further examination shows that the Line/Carter study was diachronous. This may well support Brooke's argument that diachronous and synchronous studies have varying results. As suggested by Line/Carter, the rapid decline experienced in most studies may be a result of synchronous studies (i.e. not allowing for the different size of the literature available at the time).

A report by Rouse and Rouse (1979) looked at monograph obsolescence through inter-library loan demand which represents both satisfied and unsatisfied requests.

They suggested that it was necessary to look at demand in relation to the annual growth rate of the literature or the book stock. To deal with the data they collapsed the Dewey classes to four major categories:

General

Science and Technology

Social Sciences

Arts and Humanities

They looked at actual demand as well as normalised demand. Normalised demand is defined by the ratio obtained from:<sup>76</sup>

$$\frac{\text{actual demand}}{\text{number of titles available from American Book Publishers}}$$

The analysis looked at two distributions:<sup>77</sup>

actual demand -v- publication date

normalized demand -v- publication date.

They found that correcting for the growth of the literature results in substantially longer half-lives, and that the

exponential distribution is a reasonable representation of both corrected and uncorrected data for each subject category examined.

In 1978 Gosnell looked at subject area obsolescence as opposed to individual or specific title obsolescence in college libraries. He suggested:

"The rate of obsolescence to be a property peculiar to each subject and not a variable dependent upon some other continuous variable."<sup>78</sup>

He found that larger subjects had a lower rate of obsolescence and hypothesised that lower rate subjects may accumulate more titles. (Lower rate subjects include history, english, philosophy.) He further hypothesised small subjects may be small because the obsolescence rate is high. e.g. chemistry and physics. He suggested it is necessary to include the expansion of interest in a subject, and different approaches to the subject, all of which will affect the rate of obsolescence.

Hindle's (1979) study extended the Morse Markov model use to British libraries. His study was undertaken at the General Pool Stock, Wakefield. This collection is used as a resource for inter-library loan by eleven public libraries and five university libraries. It is a

non-growth collection and at the time of the study no material had been added since 1974.

In his studies Hindle found that:

"...Morse's Markov model provided a close approximation to the pattern of use of a closed access collection serving inter-library loan requests only."<sup>79</sup>

and that:

"Sufficient data were available on old date due labels to study the pattern of use prior to being consigned to the General Pool Stock."<sup>80</sup>

The aim of the study was to obtain:

"...prediction of the probable future level of use on the Pool given that it was an obsolescent collection with no titles being added."<sup>81</sup>

and:

"...prediction of the additional usage that could be expected if the collection was dispersed to open access collections."<sup>82</sup>

Hindle found that the collection was at such a low level of average use that usage was close to the asymptotic level. He concluded that in the foreseeable future the overall level of demand would remain stable and that distribution to open-access collections would provide only a minimal increase in use.

Brown's (1980) principal aim was to show that "...obsolescence, or decrease in incidence of citation, is scientifically measureable, and can be defined by the same mathematical relation which describes radioactive decay, namely the Poisson distribution."<sup>83</sup>

Brown sampled eleven major chemical journals and plotted the average number of citations per article for the total over several years and obtained a graph which peaked at citations two years after publication of an article. She found the best fit to the average citation per article, according to a chi-squared test, was 2.9 years, and found that the half-life, allowing for the growth of the literature, was 9.3 years. She also indicated that:

"The tails of any such distribution show relatively large statistical fluctuations...[and]...account for the variety of values of the half-life calculated on the basis of the negative exponential model".<sup>84</sup>

She concluded that the Poisson distribution should give an accurate method of calculation of half-life.

Generally, what has evolved is a theory containing a number of basic measures which can be applied to individual situations that may or may not, depending on individual library needs, take into account particular user groups, the type of literature being examined and the amount of literature available.

### 1.5 Major Literature Table

The following table provides an alphabetical listing of the literature considered to have had a significant part in the development of the theory concerning collection use.

It provides a schematic view of the purpose of the research, the topic of the research and the researcher.

Researcher	Data Collection		Type of Library	Theory	Purpose Weeding	Acquisitions	Criterion
	Manual	Computer					
Brookes 1970	N/A		N/A	Serial			Utility
Brown 1980	*		Academic	Serial			Poisson Distribution
Buckland 1970	*?		Academic		*	Core Stocking	Usage
Buckland 1972	N/A		N/A	Serial			Scattering Obsolescence
Buckland 1975	*?		Academic		*	Stock Control	Variable loans, duplication policy
Chen 1972	*		Academic			*	Markov Model Analysis
Fussler/Simon 1969		*?	Research		Secondary Storage		Use in Predetermined Periods
Gosnell 1943	*?		Academic		Monographs	Monographs	Obsolescence
Hindle 1977	*		Public		Secondary Storage		Material in Store Used for Inter-library Loan
Jain 1966		*?	Academic		Secondary Storage		Relative Usage
Kent 1975		*	Research	*	*	*	Past Use Analysis
Kohut 1974	*		Academic			Serial	Citation Analysis - Obsolescence
Lawani 1975	*		N/A	Serial		Serial	Citation Analysis
Line 1970	N/A		N/A	Serial			Growth of Literature - Half Life Concept
Lister 1967		*	Academic		Secondary Storage	*	Usage
McClellan 1978	*		Public			*	Square Root Weighting
McGrath 1971	*?		Academic	Policy Develop.		*	Usage

Researcher	Data Collection		Type of Library	Theory	Purpose Weeding	Acquisitions	Criterion
	Manual	Computer					
McGrath 1978		*?	Academic			*	Subject Usage
Morse 1968	*		Academic		*		Previous Year
Rao 1973/1974	*?		N/A			Serial	Obsolescence
Rouse 1979	*		Public			ILL/Monographs	Half Life/Growth Literature
Sandison 1971	*		National	Serial			Use per Item. Types of Users, Types of Literature, Types of Search
Slote 1969 through to 1980	*?		Public		*		Past/Current Usage, Age
Trueswell 1965	*		Academic		*		Past Use
Trueswell 1966	*		Academic		*		Last Circulation Date
Trueswell 1968	*		Research	*			Items Circulated/User Identification
Trueswell 1968	*		Academic			Core Weeding	Last Circulation Date
Trueswell 1969	*		Academic			Core Definition	Last Circulation Date
Urquhart/Urquhart 1976		*?	Academic		Secondary Storage		Circulation

N/A = Not Applicable

\*? = Indicates that there is some doubt in the method of data collection

## 1.6 References

1. MORSE, Philip M. Library Effectiveness: A Systems Approach, Mass., M.I.T., 1968, p.5.
2. Ibid.
3. Ibid., p.3.
4. Ibid., p.113.
5. Ibid., p.93.
6. BRADFORD, S.C. Sources of Information of Specific Subjects, Engineering, v.137 Jan 26 1934, pp.85-86.
7. BUCKLAND, M.K. Systems Analysis of a University Library, Lancaster, University of Lancaster Library Occasional Paper No 4, 1970, p.23.
8. Ibid.
9. Ibid., p.52.
10. Ibid., p.53.
11. BUCKLAND, M.K. Book Availability and the Library User, Lond., Pergamon Press, 1975, p.57.
12. Ibid., p.137.
13. Ibid., p.138.
14. Ibid., p.139.
15. Ibid., p.136.
16. SLOTE, R. Weeding Library Collections - II, Littleton, Colorado, Libraries Unlimited Inc, 1982, p.73.
17. Ibid.
18. Ibid.
19. Ibid., p.77.
20. Ibid.
21. TRUESWELL, Richard W. Two Characteritistics of Circulation and Their Effect on the Implementation of Mechanized Circulation Control System, College & Research Libraries, v. 25 1964, p.288.

22. Ibid., p.289.
23. Ibid., p.290.
24. Ibid., p.291.
25. Ibid., p.291.
26. Ibid., p.291.
27. TRUESWELL, Richard W. A Quantitative Measure of User Circulation Requirements and its Possible Effect on Stack Thinning and Multiple Copy Determination, American Documentation, v.16(1) 1965, p.22.
28. Ibid.
29. TRUESWELL, Richard W. Determining the Optimal Number of Volumes for a Library's Core Collection, Libri, v.16 1965, p.59.
30. Ibid.
31. TRUESWELL, Richard W. User Circulation Satisfaction -v- Size of Holdings at Three Academic Libraries, College & Research Libraries, v.30 1969, p.209.
32. Ibid., p.213.
33. SLOTE, op. cit., p.79.
34. Ibid.
35. Ibid.
36. JAIN, Aridaman K. A Statistical Study of Book Use Supplemented with a Bibliography of Library Use Studies, Lafayette, Purdue University, 1967, p.87.
37. Ibid., p.88.
38. Ibid., p.124.
39. Ibid., p.23.
40. Ibid., p.125.
41. Ibid.
42. MCGRATH, William E. Relationships Between Hard/Soft, Pure/Applied, and Life/Non-Life Disciplines and Subject Book Use in a University Library, Information Processing and Management, v.124 1978, p.19.
43. Ibid.

44. Use of Library Materials. The University of Pittsburg Study. N.Y., Dekker, 1979, p.1.
45. Ibid., p.22.
46. Ibid., p.24.
47. Ibid., p.29.
48. Ibid.
49. Ibid., p.38.
50. Ibid.
51. Ibid.
52. SLOTE, op. cit., p.79.
53. Ibid., p.80.
54. Ibid.
55. Ibid.
56. Ibid.
57. Ibid., p.81, 84
58. Ibid., p.84
59. VOIGT, Melvin J. Circulation Studies Cannot Reflect Research Use, Journal of Academic Librarianship, v.5(2) May 1977, p.66.
60. Ibid.
61. Ibid.
62. MASSMAN, Virgil F. There are No Easy Solutions, Journal of Academic Librarianship, v.5(2) May 1979, p.68.
63. Ibid.
64. TRUESWELL, Richard W. Balance Library Objectives with Book Circulation, Journal of Academic Librarianship, v.5(2) May 1979, p.68.
65. Ibid.
66. Ibid., p.69.
67. Ibid.

68. SCHAD, Jasper G. Missing the Brass Ring in the Iron City, Journal of Academic Librarianship, v.5(2), May 1979, p.62.
69. MCCLELLAN, A.W. The Logistics of a Public Library Bookstock, Lond., Association of Assistant Librarians, 1978, p.8.
70. Ibid.
71. Ibid., p.52.
72. Ibid., p.53
73. Ibid., p.54.
74. Ibid., p.(xii).
75. LINE, Maurice. The Half Life of Periodical Literature Apparent and Real Obsolescence, Journal of Documentation, v.26(1) 1970, p.51
76. ROUSE, S.H. & ROUSE, William B. Analysis of Monograph Obsolescence at Two Levels of an Inter-Library Loan Network, Information Processing and Management, V.15 1979, p.221.
77. Ibid., p.222.
78. GOSNELL, Charles F. Obsolescence of Books in College Libraries, Collection Management, v.2(2) Summer 1978, p.174.
79. HINDLE, A. Markov Models of Book Obsolescence, Information Processing and Management, v.15 1979, p.18.
80. Ibid.
81. Ibid.
82. Ibid.
83. BROWN, Pauline. The Half-Life of the Chemical Literature, J.A.S.I.S. v.31(1) Jan 1980, p.62.
84. Ibid., p.63.

CHAPTER 2: DATA SOURCES AND COLLECTION

	<u>Page</u>
2.1 Introduction.	66
2.2 Population.	67
2.2.1 Sections of the Dewey Classification to be surveyed.	67
2.2.2 Areas of the collection to be used.	72
2.2.3 Material formats surveyed.	75
2.3 External Circulation.	76
2.3.1 Data collection.	76
2.3.2 Period of data collection.	79
2.3.3 Data analysis.	80
2.4 In-House Survey.	81
2.4.1 Period of data collection.	83
2.4.2 Data collection.	89
2.4.3 Data design.	100
2.4.4 Card format.	100
2.4.5 Sample data.	103

	<u>Page</u>
2.5 Reserve Room survey.	109
2.5.1 Data collection.	111
2.6 Conclusion.	112
2.7 References.	113

## 2.1 Introduction

This chapter deals with the data sources that were available for the thesis. Data was gathered from the University of New South Wales library's external circulation records, an in-house survey and a limited reserve room collection survey.

This chapter discusses the choice of ranges of the Dewey disciplines, indicates the manner in which circulation data was collected, the period of data collection, and some of the problems encountered in using full population files.

Further, there is discussion of the need for an in-house survey and the relevance of such a survey to this thesis. In-house survey data collection methods are discussed and the problems of the movement of books between sub-libraries are explained in flow chart form. To provide a simple basis for comparison (and to eliminate the criticism of not having included reserve material) a short reserve collection study was undertaken. The methodology and data collection methods are addressed in this chapter.

## 2.2 Population

### 2.2.1 Sections of the Dewey Classification to be surveyed.

The intention was to find disciplines that would be representative of one 'soft' science and one 'hard' science. It was seen as desirable to be able to provide a basis for comparison with research undertaken over the past few years into varying use patterns between subject areas, especially the work undertaken by McGrath<sub>1</sub> who developed the concept of the 'hard' and 'soft' sciences variations.

The requirements were:

- i. The size of the enrolment in the faculties likely to be using the sections of the collection to be surveyed should be sufficiently high to provide an adequate sample.
- ii. That the two "sciences" chosen would be in areas that were expected to show significant contrast in use.
- iii. That the faculties using the chosen parts of the collection provided a core of subjects with high enrolment. Despite specialisation within the faculty.

The options available included:

- a. The arts/humanities area.
- b. The economics/management area.
- c. The pure sciences area.
- d. The engineering area.
- e. The medical area.
- f. The law area.

Of these, two that met the requirements listed were the economics/management and the engineering areas.

The arts/humanities and the pure sciences were considered to be too diversified and would pose problems providing a core.

Though the enrolments in arts/pure science subjects were high, the subjects taken were largely of a service nature. This contrasted with the economics/management/engineering areas where the student programs tended to be fixed in the early years, providing concentrated use and, it was assumed, a substantial core that could be analysed.

The medical and law areas were too specialised and faculty enrolments too low to provide an acceptable sample. A further problem with these areas was the lack of machine-readable data. (The circulation systems of the medical and law sub-libraries were manual.)

It was decided that the Dewey sections to be considered in the final analysis were:

330-339            General Economics

657-658           Management

These were to represent the 'soft' sciences.

The 'hard' sciences section was:

620-629           Engineering

The chosen areas provided sufficient data for analysis and, in retrospect, if the exercise was to be redone the same areas would be selected.

In addition to the selection of Dewey sections, a class breakdown was also necessary. Initially, a manual review of the 19th edition Dewey tables was undertaken. Selection was based on the librarian's perspective of what would be considered relevant areas for review, based on collection development and use knowledge gained over several years of work with the undergraduate collection.

The type of choices made included the following classes:

330.1 Systems and theory

331.1 Labour force and market

331.8 Labour unions

332.1 Banks and banking

332.6 Investment and investments

This librarian's approach to the selection of Dewey classes was adequate. However, it did not provide a clear view of the use patterns of classes within the sections. The researcher recognised that it was likely that these were extremely specialised classes in the sections with low, research-type, use characteristics. It was also apparent that there were likely to be classes with many titles and high use that may warrant separate analysis.

To provide a more detailed view of use, computer analysis of the transaction tapes, which gave tables of title/class breakdown and transactions associated with them, was undertaken.

Some of the high moving classes that were candidates for analysis were:

331.8	606 Titles	3460 Transactions (1979/80)
-------	------------	-----------------------------

338	953	5671
-----	-----	------

338.9	792	4226
-------	-----	------

658.4	728	3421
-------	-----	------

Classes that seemed unlikely to provide meaningful results (few titles/low transactions) were:

334-334.9	34 Titles	79 Transactions (1979/80)
-----------	-----------	---------------------------

337-337.9	47	231
-----------	----	-----

339.5-339.9	35	147
-------------	----	-----

626-6216.9	1	0
------------	---	---

This computer analysis provided support for the librarian's manual approach and assisted in the development of a more viable class analysis.

### 2.2.2 Areas of the Collection to be Used.

The intention was to determine which areas of the collection would need to be examined to provide a complete analysis of use statistics.

The University of New South Wales library houses several sub-libraries within the one library building. The purpose of dividing the collection was to provide relevant and more accessible material to specific user groups.

There are two major special collections that are easily accessible in the same building. These are:

- i. Physical Sciences, covering two floors, and
- ii. Social Sciences, also covering two floors.

Both collections contain some undergraduate material for final year students in addition to research material.

Loans for all the sub-libraries are issued at a central circulation desk located on the ground floor. Borrowers are able to move freely between all floors and collections, and direct access is available to them. No restriction is placed on borrowing monographs from any of the collections, i.e. Undergraduate, Physical and Social Science libraries.

Material in the three collections is distinguished by call number prefix:

No prefix is used for undergraduate material,

'S' is the first letter of the prefix for social science items,

'P' is the first letter of the prefix for physical science items,

and within the prefix there is also a size factor incorporated with the use of:

'Q' for quarto

'F' for folio.

Location and size of materials is easily identified through this prefix system.

The two other special collections in the University of New South Wales system are:

the Law collection, and

the Biomedical collection.

Although in close proximity, neither collection was as easily accessible as the Undergraduate, Social Science or Physical Science libraries. This material has the prefixes,

'L' for law

'MB' for biomedical.

All monograph material (both undergraduate and research) can be considered as interfiled in each of these collections. There were a few exceptions arising from the use of a separate numbering system for quarto material which was in operation for a few years. In a few cases the same call number was used more than once during that period. The errors arising from discounting this were considered to be acceptable and consistent with certain other errors involved in the data processing. This simplification greatly assisted the processing.

All material is available for loan (except for 'reserve' items and works held in reference collections) and the prefixes are recorded as part of the call number field in the loan record. This unique prefixing enables external circulation comparisons to be made for use of the same work from the different sub-libraries and duplicate copies of undergraduate material in the special libraries can be easily recognised.

### 2.2.3 Material formats surveyed

Consistent with research undertaken by Kent<sub>2</sub> and others, the aim of this study was to examine the use by undergraduates of monographs.

Monographs were considered to be important in their own right and a comprehensive study was seen as essential. The questions of 'core' and the relationship between user groups and 'core' in monograph collections was the primary concern of this thesis, with the ultimate intention of providing data for improved collections for teaching, and more cost-effective use of acquisition funds.

A monograph study in the University of New South Wales library provided a large exercise. The addition of other formats, e.g. serials and audio-visual material, would have made the research unworkable because of the size.

The serial collection requires separate study. Not all serials are available for loan, consequently the methodology used to examine monographs might not be applicable. Further, the way in which this part of the collection is used cannot be expected to be necessarily similar to that of monographs.

## 2.3      External Circulation

### 2.3.1    Data Collection

At the time of circulation automation, all copies of monograph material in the undergraduate collection had a machine-readable 80 column card inserted. All subsequent material had cards inserted at the cataloguing stage.

With the transfer of material to the special libraries in 1975, machine-readable cards were inserted with the library's prefix incorporated into the call number.

Upon completion of the loan transaction, the data is transferred to a statistics file held on magnetic tape. The statistics record includes the following data:

- \*    Call number (prefix plus Dewey classification number) together with book, cutter and copy numbers.
- \*    The date the item was borrowed.
- \*    The date the item was returned.
- \*    Borrower identification -- a unique number indicating year of enrolment for undergraduates with special indicators for postgraduate and academic staff.

Schad, in his criticism of Kent, discussed the problems of errors in loan records. To establish whether this was to be a problem with this research, an analysis of the error reports associated with the circulation system was undertaken. It was found that it was highly unlikely that there was any significant error in the records maintained for loan transactions, apart from the occasional misread of either book or borrower identification. From all indications, the error rate was very low.

Note that what goes on the statistics file is the record of the completed transaction. If there was an error made at the time of borrowing in, say, the call number, then it will not be possible to cancel that loan unless the original data is corrected. The date information is not encoded on a per item basis but is set up by the computer. The IPL incorporates checks that should make the dates very reliable. The only field that may contain errors is the borrower identification number. The borrower card is made of strong material so errors there are not likely. If they do occur then the only digits of significance to this research are the first two digits of the identification number. In short, one may expect that the data as it went onto the tape had a very low error level.

In the initial analysis of the data, items that were on loan to:

- \* Open reserve (reading room material in high use, usually text and reference material for undergraduate students),
- \* binding section, for repairs and rebinding,
- \* other libraries (outside the University system),

were removed from the data.

The removal of the data was possible since in all cases the material on loan to the three sections indicated had a special borrower identification number. Elimination of these transactions (as part of the data) removed any distortion they may have caused.

Material in Open Reserve is assumed to have high (internal) use. The material does not figure in (external) circulation analysis since it is not borrowed in the sense in which that term is used in circulation studies. The only way in which the reserve copy appears is on return from reserve to the open stacks; the "loan" is then over and the historical record is written.

It is difficult to resolve how best to incorporate internal reserve "loans" and external loans into the analysis. Each represents a different use and it may not be possible to add the figures directly without producing misleading results. The reasons for the analysis may well be different in the two cases so only one of the two figures may be relevant. For example, a decision has to be made whether reserve items should be returned to the open stacks. Here both internal and external use are relevant though they are not directly added. An absence of reserve use (internal circulation) coupled with some external borrowing conveys a picture where both numbers are relevant. Such use may imply that copies are not used in reserve because adequate numbers are available in open access. If only the reserve (internal) use was examined it would be assumed that the item was not a high mover. Effectively, there is a range of problems where one really needs a multi-dimensional description of use.

This thesis did not look at such issues, though the author had been involved with them. The paper written on the Macquarie weeding experiment discusses them more fully.

### 2.3.2 Period of data collection

To establish use patterns, the completed transaction records were analysed for the following periods:

1971, 1972, 1973, 1976, 1978-80.

Problems experienced with the data files are discussed in detail in Chapter 3.2.1.

However, since the data analysed represented the total external circulation of monographs<sub>3</sub> from the collection, it was considered that the results obtained would reflect use in the three years omitted. (There were, however, considerable difficulties in the longitudinal analysis due to absence of this data.)

### 2.3.3 Data analysis

The analysis of data required flexibility and had to be handled on a trial and error basis. Analysis options and aspects of computer processing are dealt with in detail in Chapter 3 et seq.

## 2.4 In-house Survey

In examining the external circulation data to determine criteria for weeding and the relegation of material and when looking at the 'core' that may be applicable in undergraduate collections, it was apparent that in-house use may need to be considered.

Kent<sub>4</sub> remarks that this group is mutually exclusive of external circulation since material used in-house is not available at that time for (external) borrowing. (In the University of New South Wales delays in reshelving, as discussed later in this chapter, would need to be considered when assessing multiple use of items.)

The importance of an in-house survey is that it provides some insight into the total use patterns. It gives a feeling for what may be excluded if external circulation was all that was surveyed.

In this thesis, in-house use has been examined from a particular perspective. Here we provide the rudimentary development of a cluster theory analysis. This may be continued at a later date to look more closely into the specific behaviour patterns of in-house use. The results of this analysis are contained in Chapter 4.6.

This section looks at the data collection aspects of the in-house survey. In particular we examine the decisions made at the time on the length of survey and the areas of classification which were surveyed. Here are outlined many of the problems experienced and the implications on the interpretation of the results that arose from the procedures.

The criticism levelled at the Kent Study by Borkowski<sub>5</sub> of the validity of a small in-house sample was not overcome in this study. (Borkowski's criticism was that a 30-day in-house sample and its extrapolation as the basis for a seven-year study was flawed.)

Some of the literature on in-house surveys expressed concern that senior faculty and librarians reshelve material after use, as part of the consultative and browsing aspects of research. Further, these users are likely to look at obscure material that does not appear on circulation records. Such behaviour, it was felt, would distort both in-house and external use patterns.

This study was analysing use of material located for use by undergraduates and it seemed unlikely that such use would distort the results of the in-house analysis.

#### 2.4.1 Period of data collection

The intention was to monitor activity during a normal teaching period. This would mean that there could be contributions to in-house use due to several user categories, including undergraduates, academics and postgraduates.

The period selected was the three weeks from September 30, to October 21, 1978. This was chosen after examining the previous two years' circulation statistics. It was considered that this period would reflect use of material with fewer distortions (e.g. examinations, low or heavy assignment loads, semester breaks, etc.) than other periods.

The options, available for the sampling interval, were:

- i. Monitor the full semester,
- ii. within the teaching period
  - a. at the start of semester
  - b. at the end of semester
  - c. near the mid-semester break
  - d. within the semester and away from the break,  
beginning and end of the semester,
- iii. before and after the teaching weeks,
- iv. over a period that included the mid-semester break.

The option chosen was (ii) (d).

Option (i) was not workable because of the volume of data that would be generated. Options (iii) and (iv) were rejected on the basis of not being able to monitor a teaching demand (Option (iv) may be regarded as "teaching" but there may be atypical influences of assignments, etc.)

There was concern with Option (ii) (c) due to the effects of the break and it, too, was rejected.

It is the experience of reference and circulation librarians that the pattern of use alters towards the end of the session. The presence of examinations leads to a wider use in terms of the topics considered and yet a more concentrated use of particular material than normal. This was the reason that Option (ii) (b) was not chosen.

At the start of semester, use is normally low since assignments are seldom due in the first few weeks of semester. The use of Option (ii) (a) would have produced a distorted sample in that there was unlikely to be an assignment component and it was also eliminated.

It may be argued that this line could be used to suggest that no part of the semester was "normal". This is a reasonable proposition. A simple view of "normal" may be non-sustainable. Rather, use may have a pattern over time and it may be only realistic to look at one feature of that pattern. This may be a topic for

subsequent study and attention here will be directed to a simplistically defined "normal".

A requirement in the sampling was for a sample size of at least 2000. This figure was chosen as it was sufficiently large, both to be representative of the population and to permit estimates that were valid in the development of the cluster analysis. The sample was to include all eligible material, i.e. all material in the Dewey classes chosen for study, during the survey period.

A pilot survey was carried out over three days prior to the main survey. It yielded 100 monographs per day from the chosen Dewey classes.

Based on the pilot study, in order to obtain the 2000 items for the sample, the study was run over three weeks. The time period chosen for the sample is considered to have achieved what was required.

Additional information that may have proved relevant in determining the most suitable interval for sampling, was knowledge of the "due dates" of assignments from large enrolment subjects that may have been using the areas of the collection being surveyed.

This survey, of course, is unable to indicate whether any section of the collection sampled was experiencing unusual

use resulting from assignments or to make assessments of use patterns before and after the survey period.

The unique prefixing, as discussed in 2.2.2, enabled the tagging of material in two ways:

Firstly, by use of 'home' floors, i.e. material being used on the floors on which it was shelved, and,

secondly, by use of material away from its 'home' floors, i.e. material from a specific collection that had moved to areas other than those on which it was shelved.

This helped to determine whether there was movement across collections. Relevant questions were whether students sought high demand material in other parts of the building if unavailable on the 'home' floor, and also whether material used in-house travelled across collections. In the latter case it may have been used on a floor away from its 'home' locations by students looking for quiet spots to study.

The options available for the in-house study included:

- (i) All research collections only,
- (ii) one research collection and undergraduate collection,

- (iii) two research collections and undergraduate in the same building,
- (iv) undergraduate only,
- (v) undergraduate, research collections in the building, and Law and Biomedical collections,
- (vi) Law and Biomedical collections and undergraduate.

(i) would not be consistent with the aim of the research, which was primarily to look at undergraduate material of interest to undergraduates.

(ii) was eliminated. The physical layout of the University of New South Wales library system provides easy movement through two research collections, coupled with the fact that the choice of one research collection (either Social Sciences or Physical Sciences) would not give an adequate picture of the areas chosen since a choice of one would eliminate either the 'hard' sciences area or the 'soft' sciences area.

(iv) would miss final year undergraduate courses whose material was housed in the research collections, and would make the survey incomplete. This option was also rejected.

(v) and (vi) include the two special collections that were not as easily accessible.

It was assumed that there would be only casual use in these libraries of the parts of the collection under review.

Although both the Law and Biomedical libraries have a core of material that covers all the classes of interest to this thesis, they only provide basic support material related to the specialist libraries.

Although the undergraduate, social sciences and physical sciences collections are physically separated, it was considered that there was likely to be interchangeable use between them. That is, if a work is held in more than one sub-collection, then the user who does not find what he is looking for in one place will look elsewhere. For this to occur, it is necessary that the catalogue entry indicates the holdings of the various sub-libraries. This is so at the University of New South Wales.

Option (iii) was considered the most valid and was consequently chosen.

As discussed in Section 2.4.1, monographs were the material format analysed in this work.

The in-house study was to be supportive of the external use study. An in-house survey that included serials would have introduced a host of new issues that could not easily be explored. (For example, in the preparation of assignments, journal articles may have a place in Economics that they do not occupy in the handling of tutorials or the review of lecture material.)

#### 2.4.2 Data Collection

A number of issues were critical in determining both the best method and the period for data collection. Opening hours, staffing levels and varying shelving procedures in the sub-libraries were important. The movement of monographs across floors required the researcher to make additional adjustments to the data collection. These issues are discussed in detail in this section.

#### LIBRARY HOURS

Over the survey period, the library was open in the hours:

Monday-Thursday	8.30 am to 10 pm
Friday	8.30 am to 6 pm
Weekends	12 noon to 6 pm

## STAFF

The staff organisation of the Readers Services' Division in 1978 is shown on the organisation chart (Figure 2.1).

Attendants collect and discharge returned material for the undergraduate collection. These staff are also responsible for the collection and shelving of material used within the library.

The attendants perform these functions on a roster basis in order to fulfil the other requirements of their jobs, e.g. security duties, maintenance of equipment, etc.

The Social Sciences library also uses its attendants to collect books used within that library, to sort books (both those used within the library and those returned from loan whose location is the Social Sciences and Humanities library), and to shelve this material.

In both the Undergraduate and Social Sciences and Humanities library at the time of the survey, the full-time attendant staff was supplemented by part-time student assistants who helped in the tasks of discharging, collection, sorting and shelving.

The Physical Sciences library employed two attendants daily, one on each level of the library, for the

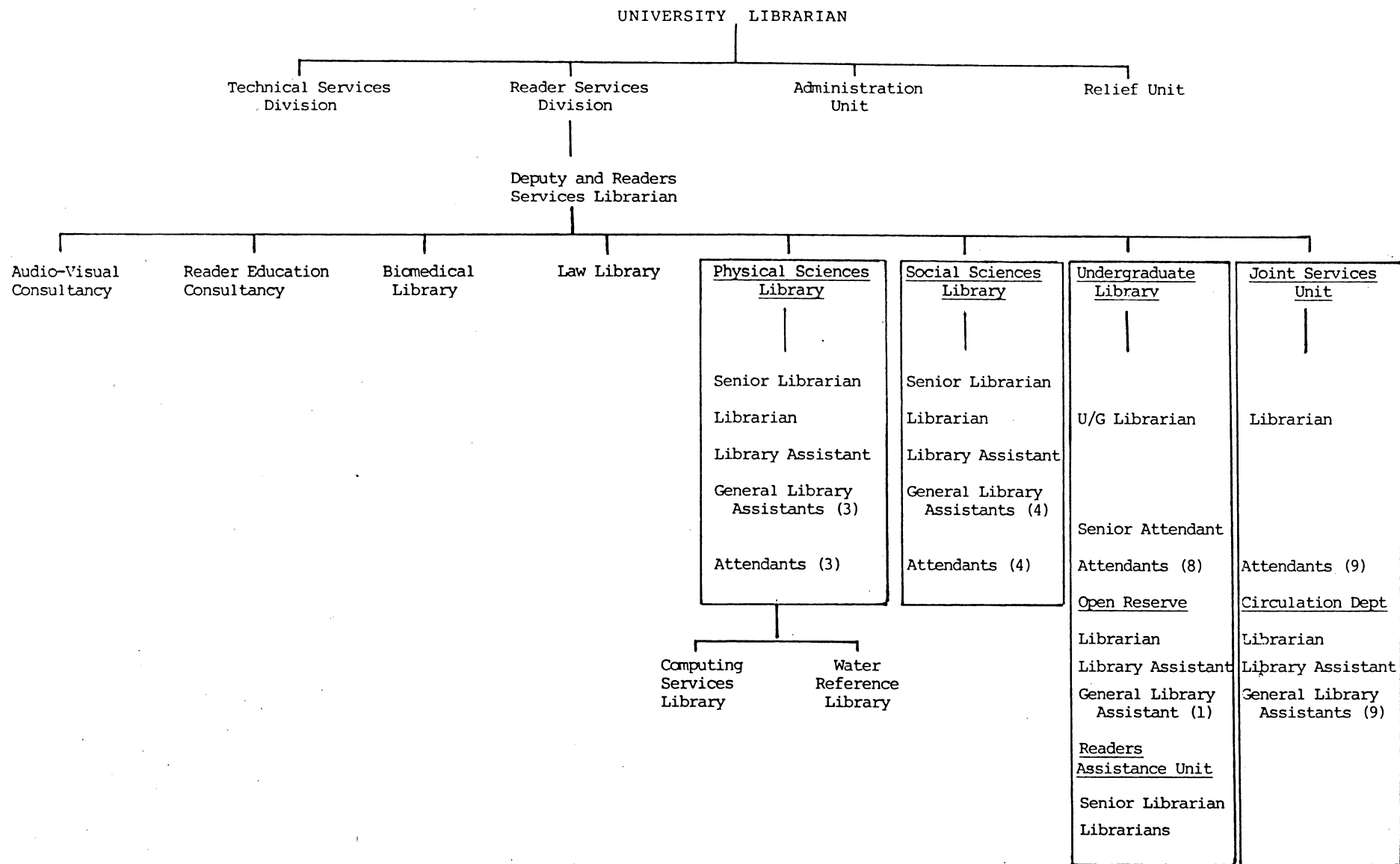


Figure 2.1 Organisational Chart of the University of New South Wales Library as at January 1978  
 Note that it is not representative of all Library staff, but it does show staff

collection, sorting and shelving of material. There was no student support.

Returned material from the Undergraduate library or the Social Sciences and Humanities library was sorted into classification order in the basement and returned to its appropriate location. Unsorted material was sent back to the Physical Sciences library.

(1) Undergraduate Library

Material received from the basement was shelved immediately.

Material that had been used within the library was collected and placed in the sorting room where it was sorted into classification order and placed on trolleys for shelving.

To determine what material was being used in-house, it was necessary to distinguish between monographs which had been on loan and those that had appeared in the sorting room as part of the general in-house use.

For the survey, in-house trolleys were marked by the attendants, indicating that the material was in-house use only, and also the date that the material was placed on the trolley.

Eligible survey material was taken from:

- (i) trolleys of sorted monographs which had been marked by the library attendants as having been collected in-house,  
and,
- (ii) unsorted monographs still located in the sorting room.

To avoid including monographs that had been on loan and which may have been missed in the initial checks, an added check was done. The last cancellation stamp on the date due slip for all monographs was checked. If the cancelled stamp was either the same day or the previous day's (the sample day), it was assumed that the book had been on loan and it was not included in the sample. (It was necessary to include the previous day because of time delays involved in getting books from the discharge point in the basement to the relevant floor for reshelving.)

To avoid duplication of the sample at a later date, trolleys were marked with yellow spots showing the date and the floor on which they had been previously checked. Material is housed over several floors in the Undergraduate library and it was not uncommon for the checked trolleys to appear on another floor at a later date.

The material that was to be included from these levels was all monographs with no prefix, i.e. undergraduate material in the Dewey classes

330-339, 620-629, and 657-658

Any prefixed material from either the Social Sciences and Humanities library or the Physical Sciences library collected on these floors was recorded as a use on the floor on which it was found.

This was necessary to determine whether there were significant use patterns across libraries.

#### Movement of Undergraduate Library Material (loan and in-house)

The following flow chart (Figure 2.2) indicates the movement of material in the Undergraduate library.

Monographs could be picked up on Levels 3-7. Consequently, material needed to be sorted and sent to the appropriate home level.

The shelving procedures may have led to sample monographs being counted three times:

# Movement of Undergraduate Materials (Loan and In-House)

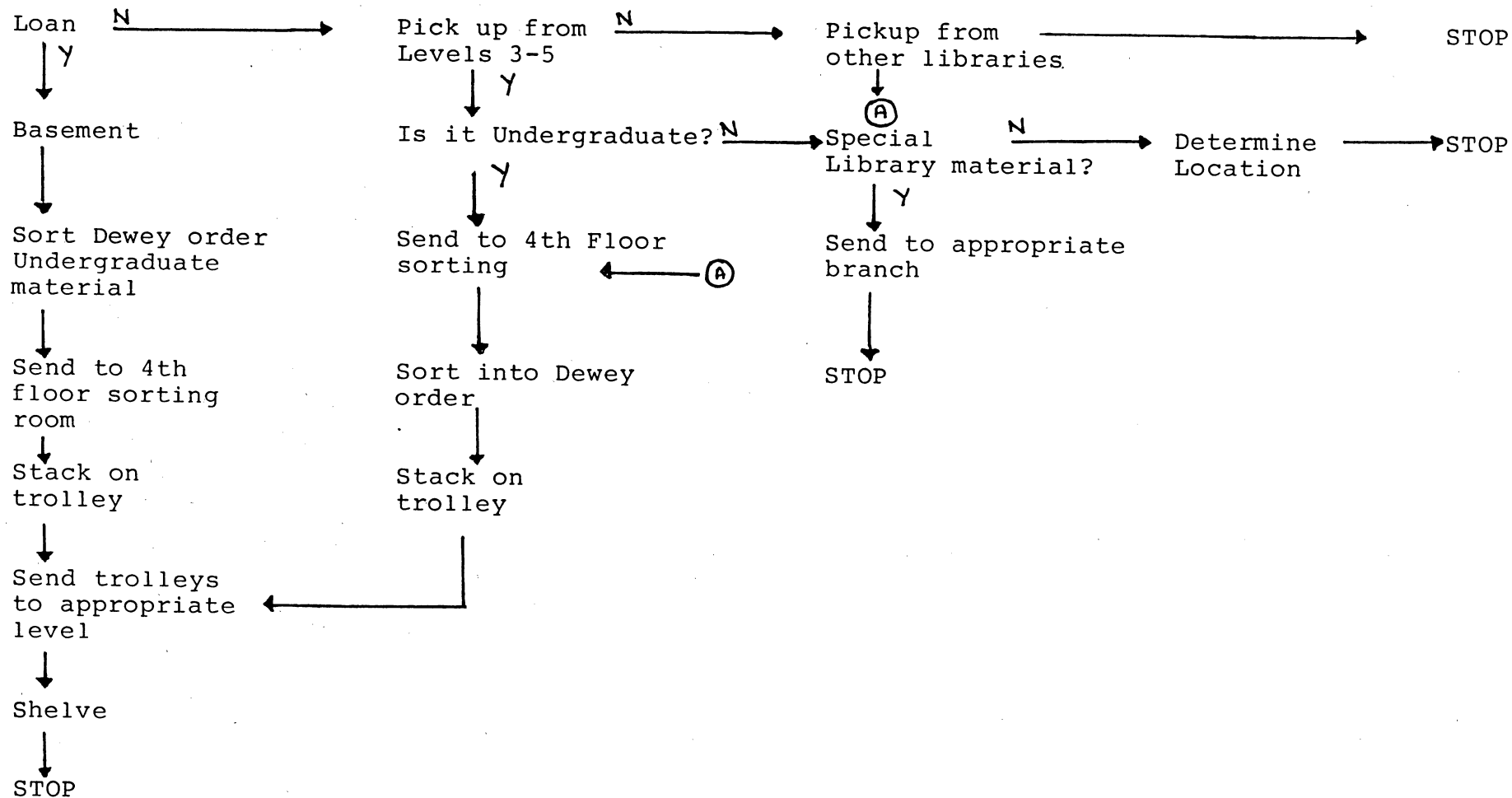


Figure 2.2

- a. in the sorting room,
- b. waiting outside the sorting room to be sent to the home level,
- c. On the home level awaiting shelving.

This problem was overcome by using the methodology discussed earlier.

(ii) Social Sciences and Humanities Library

Material in the Social Sciences and Humanities library was also surveyed and it was necessary to distinguish material that had been on loan from that used in-house.

Material that was returned from loan was interfiled in the sorting room with material that had been collected in-house, thus making it difficult to easily distinguish material eligible for the sample.

To determine which monographs had been used in-house, every monograph in the sorting room and any sorted trolleys waiting for reshelving had to be checked.

This was done by checking the cancellation stamp on each date due slip and applying the same criteria as were used in the Undergraduate library to determine whether a monograph would be included in the sample.

At the time of the survey there was a backlog of shelving which caused some duplication of the sample.

Movement of Social Sciences and Humanities Library Material  
(loan and in-house)

The following flow chart (Figure 2.3) indicates the movement of material in the Social Sciences and Humanities library.

As in the Undergraduate library shelving procedures may have led to sample monographs being counted three times,

- (i) in the sorting room,
- (ii) waiting outside on trolleys,
- (iii) on the home level awaiting shelving.

This duplication occurred despite the marking with yellow spots, giving the date and floor. Material included:

All social science, physical science and undergraduate material found in the Dewey classes being reviewed.

# Movement of Social Sciences and Humanities Library Material (Loan and In-House)

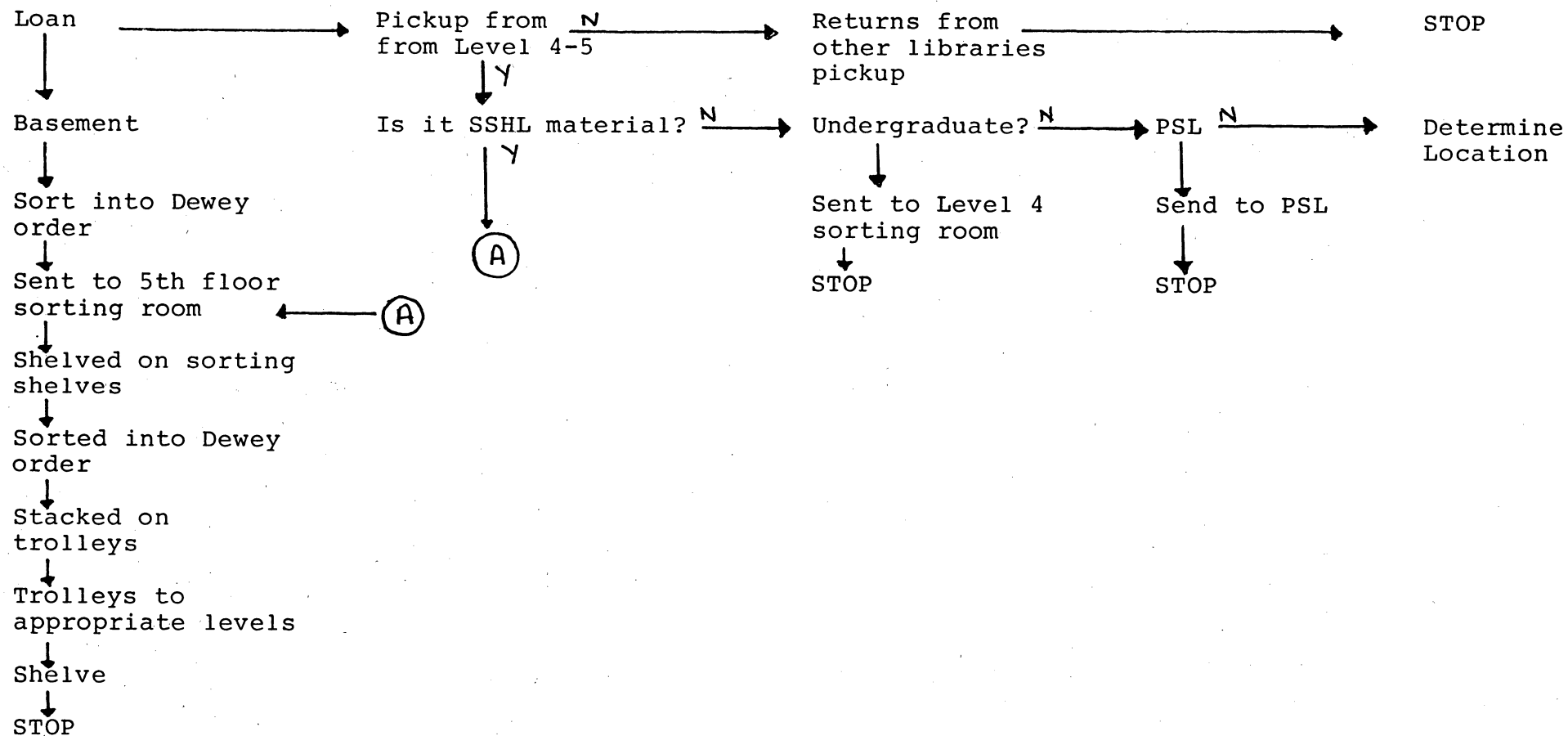


Figure 2.3.

Material is collected on Levels 4 and 5 and sent to appropriate home floor.

(iii) Physical Sciences Library

Returned loan material sent up from the basement is kept separate from the material picked up from in-house use.

There is no sorting room facility in this library, making it necessary to have designated deposit points for returned material on Floors 6 and 7.

On Floor 7 the cancellation stamp was checked and the determination of whether the item had been used in-house was made by applying the same criteria as those used in the other two libraries.

The sample collected was small as very little material gathered at the deposit point on this floor and clearing and reshelving were done daily.

Floor 6 posed some problems. Material at the deposit point on this floor was normally shelved by 10 am.

To simplify the survey task for the attendant all material used in-house was recorded. At a later date, serials and monographs not belonging to the sample Dewey classes were eliminated.

Material belonging to the Physical Sciences library, but found on other floors, was added to the sample on those

floors. This was necessary because shelving on Levels 6 and 7 was done promptly each day.

### Movement of Physical Sciences Material

The following flow chart (Figure 2.4) indicates the movement of material in the Physical Sciences library.

#### 2.4.3 Data Design

To assist in the development of the cluster analysis, it was necessary to include:

1. full call number, including copy number and prefix  
(prefix indicates special library location),
2. day of year surveyed,
3. floor on which the material was found,
4. Julian day and month surveyed.

This data was collected manually and each transaction was recorded separately on an 80 column card.

#### 2.4.4 Card Format

For processing purposes, the transactions were recorded on 80 column cards in the following format:

Column 1-34	expanded call number
Column 73-75	day of year surveyed
Column 76	floor on which monograph found
Column 77 and 78	Julian day surveyed
Column 79 and 80	Julian month in which surveyed.

The sample was to be collected daily over the three week period, including weekends.

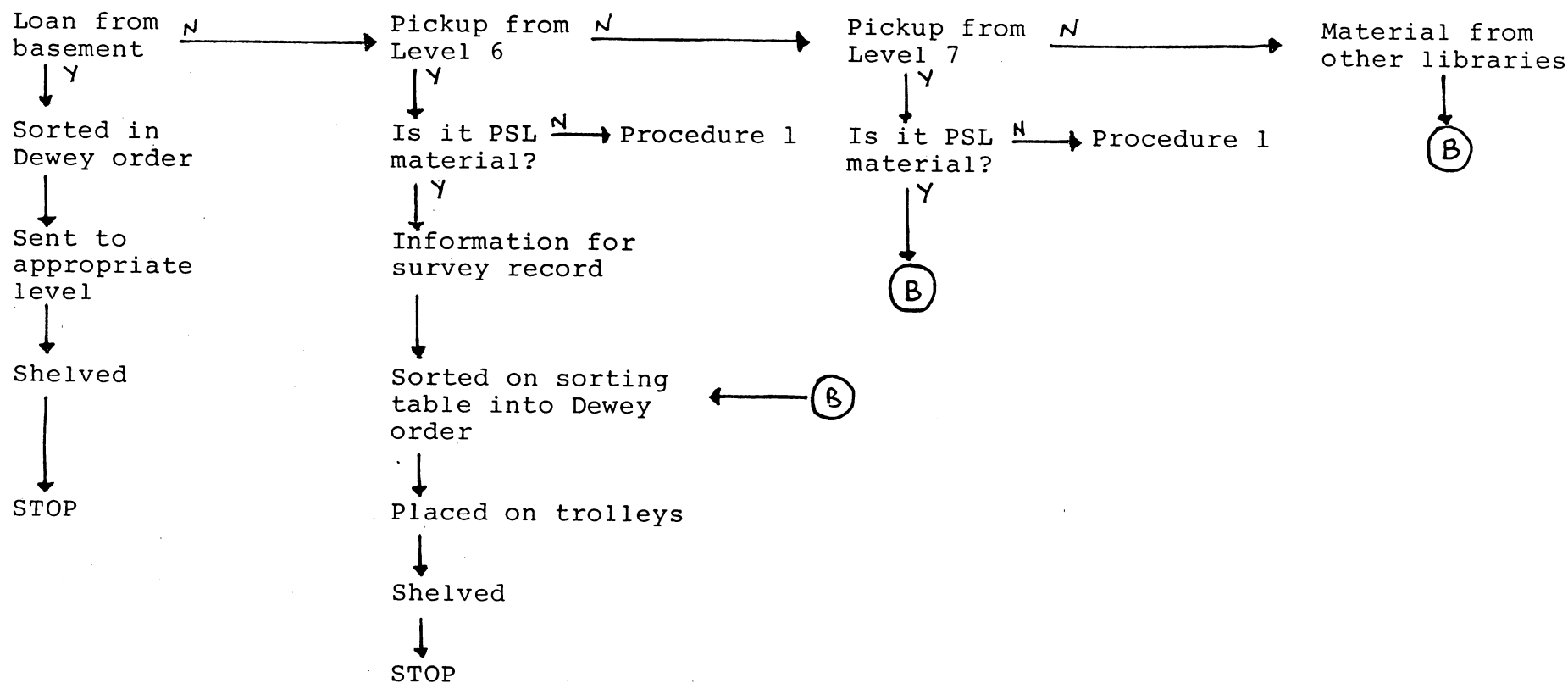
All eligible material used daily in-house was included. Early Friday night closing necessitated collection of the sample at opening time on Saturday. This was deemed equivalent since no material would have been collected or shelved between closing time Friday and opening time Saturday.

Sunday surveys were not undertaken until late afternoon to allow material from Saturday to be collected and sorted. This was necessary as there were fewer staff to do these tasks on the weekends.

Flows of material through the special libraries differed and different sampling techniques were required.

Floor 6 material was recorded in the morning rather than the afternoon since the flow of material would have made this sample invalid. (See Figure 2.4.)

# Movement of Physical Sciences Material (Loan and In-House)



## Procedure 1

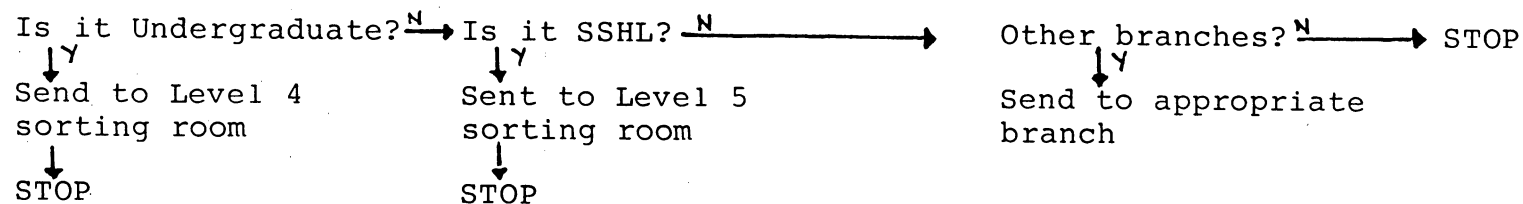


Figure 2.4

#### 2.4.5 Sample Data

The material flow, discussed earlier, created some problems. The Undergraduate library and the Social Sciences and Humanities library during the pilot study and the sample period, were experiencing some delays in the reshelving of material.

The undergraduate material, however, was separated into loans and in-house use. Dating the in-house use trolleys avoided the inclusion of the same material twice as it moved from the receiving floor (Level 4) to the location floor (Levels 5, 4, or 3).

The problem of overflow reshelving in the Undergraduate library occurred daily but was more pronounced on weekends. Early (6 pm) Friday night closing meant that a proportion of the weekly shelving was not completed and flowed into the weekends.

The Social Sciences and Humanities library was a more serious problem. The material was not divided into two categories and was required all to pass through the sorting room (Level 5) for resorting. This created reshelving delays of up to four days.

All material for reshelving was located outside the Level 5 sorting room until the reshelvers collected trolleys for

shelving on the location floor (Levels 4 and 5).

The Social Sciences reshelving problem was not limited to particular days and for the whole sample period it occurred continually. Distortion through duplication in the sample of this collection was unavoidable unless certain adjustments were made.

The speed of reshelving in the Physical Sciences library eliminated the problems experienced in the other two libraries.

The pilot survey revealed that duplication would occur for other reasons apart from the major reshelving problem discussed above.

These included:

1. The movement of sorted material to its location floor on the same day would, on first assessment, be considered two uses.
2. There was evidence that some material was being found on consecutive days on the same floor. Genuine two-use material would have to be separated from reshelving delays.

3. Material on a non-location floor on Day 1 and on its location floor on Day 2 could also have been perceived as two uses. Some decision would be necessary, based on the flow of materials.

The following adjustments were to be made to the data from the in-house survey to eliminate any duplication that may have occurred from the factors in 1-3 above.

- a. Material found on a non-location floor and on its location floor on the same day was assumed to be in the process of normal reshelving. In this case, location floor data would be eliminated.
- b. The need to determine genuine consecutive day use of certain material could only be achieved by eliminating material not in that category. It was felt that the introduction of a yellow spot on each trolley (with a date indicating that that material had already been included in the sample) would overcome the double counting of material subject to reshelving delays.
- c. Material moving from a non-location floor to its location floor on consecutive days was to be considered as part of the normal reshelving process. Location floor data was to be eliminated and the material was to be recorded as having one use.

- d. The special problems in the Social Sciences and Humanities library required additional adjustments. If materials appeared on both the sorting floor (Level 5) and the location floor up to or including two days later, the location floor use was not included.

Material appearing more than once, up to or including four days later on the sorting floor, was considered as one use only. This type of delay was assumed to be a reshelving problem.

The flow of material not only created the duplication problems listed above but a number of other combinations of movement required decisions to be made regarding the number of times material was to be considered as having been used.

- e. Material found on a non-location floor, on a second non-location floor, and on its location floor, all on Day 2, on first appearance, indicated three uses.

Use of this type was seen to have been used only twice. The movement from a non-location floor on Day 2 to the location floor in Day 2 was seen to be part of the reshelving process and not as another use.

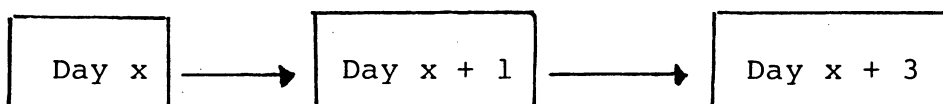
- f. Material found on its location floor on Day 1 and on a non-location floor on Day 2, indicating movement away from its location floor, showed more than one use.

Use of this type indicated movement away from the location floor. This was perceived as being extra use and was, therefore, to be recorded as having been used twice.

The different shelving procedures on Level 6 (Physical Sciences library) created further combinations of material movement requiring some decisions to be made regarding use.

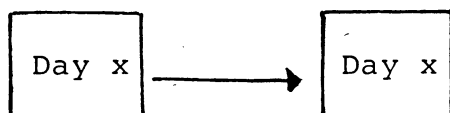
These procedures included:

- g. Floor 6                      Non Location                      Location Floor 6



and

- h. Floor 6                      Non Location



Use of the type discussed in (g) was seen to be three uses and (h) was to be treated as having been used twice.

To avoid excessive handling, the adjustments were to be made after all the data was collected.

This was done by sorting the data in such a way that combinations would be easily observed.

The sorts were:

1. Call number as the primary sort and date as the secondary sort. The call number prefix was included in this sort.
2. Call number as the primary sort and date as the secondary sort. However, in this case the call number prefix was excluded, permitting consecutive day duplication to be easily determined.

## 2.5 Reserve Room Survey

Material placed in reserve is expected to be high-use text and reference material. Since this thesis is primarily concerned with the concept of 'core' with the view to addressing such issues as multiple copy purchase, stock relegation, secondary storage and weeding, any analysis that did not include at least a brief overview of this high-use area would be incomplete. It is by no means an exhaustive study but a three-week sample was seen as providing a reasonable basis for comparison. The results of this reserve survey are given in Chapter 4.7.

Material placed in this collection is captured by the circulation system, providing a machine-readable record of material on loan to the reserve collection. At the time of the study, use within the reserve collection was a manual process.

Items in reserve were charged out for a four hour period, renewals were possible provided other students had not reserved the material, and use was limited to within the library, although some material could be borrowed after 9.45 pm for overnight use. Material in reserve included monographs, reprints of journal articles and personal copies of books and journals belonging to lecturers (usually out-of-print material, or material not available in the library).

The reserve room operated on an open stack basis for monographs. Students obtained the monograph they required from the shelf, took the item to the desk where it was date stamped in the book, and were given a time card (punched on a Bundy machine) that showed when it should be returned. No other record of the loan was maintained. This made it difficult for the research study to collect data at the time of the loan.

The Kent Study<sub>6</sub> looked at loans to the reserve room collection but did not capture data for loans within reserve. Since this area is considered high-use, it was thought essential to provide at least a limited comparison of reserve use and external circulation.

As indicated in 2.4 the In-house Survey, the period chosen was one which would indicate normal teaching activity and, to make a valid comparison, the reserve survey was conducted during the three weeks selected for the in-house survey, September 30 to October 21, 1978.

Material borrowed from 'Open Reserve' was date-stamped every time the book was lent, making it possible to collect loan statistics retrospectively. Data could therefore be collected without time constraints.

### 2.5.1 Data Collection

Data collection was in the reserve book area of the library.

All monographs in the chosen disciplines were checked for date-due stamps from the period September 30 to October 21, 1978. The call number of the book was recorded manually under date, and where more than one loan had occurred on a volume in a day, the number of times issued was also recorded.

This data was then transferred to 80 column cards to provide machine-readable records. This would allow easier manipulation of the data for comparison of reserve room use curves, with in-house and external circulation use curves.

The data recorded were:

- . a consecutive control number for identification of each record,
- . the call number of the volume,
- . the date borrowed,
- . an open reserve indicator (for ease of processing).

Data was sorted by call number as the primary sort and date as the secondary sort. Call number prefix was excluded in this sort.

## 2.6 Conclusion

This chapter has dealt with the data sources and the collection problems experienced in the establishment of relevant data for external use, in-house use and reserve collection surveys. Some of the problems discussed were unique to the University of New South Wales library and could not be avoided. The factors of duplication and shelving delays can cause major problems in this type of survey but control of these can be difficult.

Future work, in the area of in-house surveys, would require more control over these problems, especially if a detailed long-term study was undertaken.

The problems of file structures and processing of data are dealt with in the following chapters.

## 2.7 References

1. MCGRATH, William E. Relationships between Hard/Soft, Pure/Applied and Life/Non-Life Disciplines and Subject Book Use in a University Library. Information Processing and Management, v.124 1978, p.19.
2. Use of Library Materials. University of Pittsburg Study. N.Y., Dekker, 1979.
3. 'Monograph' is used to describe books in the narrower sense, i.e. publications that are not serials.
4. Use of Library Materials. op. cit., p. 27.
5. BORKOWSKI, C. Report on the Kent Study of Library Use: A University of Pennsylvania Reply. Library Acquisitions: Practice & Theory V.3 1979, p.134-135.
6. Use of Library Materials. op. cit., p.34.

CHAPTER 3: FILES AND COMPUTER PROCESSING

	<u>Page</u>
3.1 Introduction.	115
3.2 Data Files.	117
3.2.1 Circulation Statistical Data Files.	117
3.2.2 Analysis Options.	119
3.3 Aspects of Computer Processing.	125
3.3.1 Processor Requirements.	125
3.3.2 File Development.	128
3.3.3 Ingres Processing.	133
3.3.4 Circulation Relations.	135
3.4 Conclusions.	139
3.5 References.	140
Appendix 3.1 Relational Data Base Structure.	141
Appendix 3.2 Ingres Commands.	148
References to Appendices.	156

### 3.1 Introduction

This chapter deals with the circulation statistical files that were available for analysis and discusses some of the problems associated with these files.

Discussion of the available processing options gives a brief overview of the packages used and looks in some detail at issues associated with the choice of a database management system.

Section 3.3., "Aspects of Computer Processing", introduces several important issues. Included is a short discussion on processor requirements and the problems encountered in this research in dealing with several computer installations. There is lengthy discussion on "file development". This looks at the procedures used to create the files required for the analysis of data.

The section on "Ingres Processing" discusses the analysis and the form of the "arrays" that enable comparative review of use by academics, postgraduates and undergraduates for the years under review.

Two appendices have been included. The first is a review of relational database structures. This should assist understanding the use of these file structures and it will

provide some explanation of jargon. The second lists and describes functions of the Quel (query language) commands used with the database management system Ingres.

Overall, the chapter is designed to give an overview of computer processing used in this thesis.

### 3.2 Data Files

#### 3.2.1 Circulation Statistical Data Files

The files recording historical transaction detail provided the basic data. These were generated as a byproduct of the circulation loan recording and discharging system.

Data to establish a loan record comes from three sources. A machine-readable card held in a pocket in the book contains unique identification of the book; from the user's identification card comes unique patron information; and from the system comes the dates of borrowing and when the book is due for return.

After the item has been returned the data associated with the loan is written to an archive tape. The data recorded included:

1. call number,
2. borrower's identification number,
3. a shortened version of author/title for identification purposes (for the user not the computer),
4. date borrowed, date due and date returned.

The analysis of the thesis, like the Kent study, intended to focus on the entire population of items on loan, since complete historical data was thought to be available.

Some of the files, however, had undergone formatting changes in 1976 and 1977 which made that data unusable. The date part of the record had been corrupted by a "tidying" operation and was not reliable.

Some of the tapes had not been properly stored and, over a period of years, they had become unreadable. In other cases there were large numbers of blocks that were unreadable. Thus, a read of the good blocks could have caused errors in the statistics if there were bad blocks in the Dewey ranges of interest. Further, the tape handling routines available on the VAX 780 were primitive and recovery procedures that could have been practicable on sophisticated drives were not possible.

In short, at the start of the project it appeared as though a definitive longitudinal study was possible. But a brief skirmish with the tapes showed that the 10-year period had many gaps. Accordingly the emphasis had to be shifted to a series of one-year analyses.

Wherever possible full statistical files were used to produce frequency distributions and for establishing comparative analyses of various parts of the collection; these will be discussed further in Section 3.3.

The circulation programs that existed for analysing these files provided only the capability of borrower-type

breakdowns with overall totals of the number of loans on a user category basis. The type of analysis this thesis required was more detailed.

### 3.2.2. Analysis Options

Statistical packages such as SPSS.X and BDMP statistical software were considered and compared with use of high-level query languages on a database management system. In all cases data would need to be downloaded from the tapes. SPSS.X and BDMP supported a range of sophisticated analysis procedures, e.g. regression analysis, cross-tabulations for various characteristics,  $\chi^2$  tests and t-tests.

However, the type of questions that this thesis was attempting to answer were simple. Examples are 'What portion of the collection accounted for what percentage of use? What titles are high movers? How do various areas of the Dewey disciplines compare?' These require only simple processing of the 'count' and 'Boolean selection' type and do not need the sophistication of SPSS.X or BDMP. Both these packages were capable of doing the task but their use required some detailed record formatting and specification.

The library's statistical file format and application programs were data dependent. This means that the data

was organised and accessed according to the requirements of the circulation system application and the data organisation and access techniques were built into the application logic. For example, sort order was important and applications relied on the files being in some predetermined form that was specified on a byte basis. This format and application method was efficient for the type of analysis and record keeping done by a circulation system; indeed the efficiency that results from it is important in transaction handling. However, it was apparent that for the variety of purposes of this thesis a preferable approach would be one that did not presuppose sequencing, indexing or the like. It was useful to have research programs that were not dependent on the storage structure or access strategies. Note that this opened the application to the use of high-level query languages which do not require particular storage arrangements. Following this line, it was natural to look to use of a database management system.

The prime advantages of such systems include integration and support of the amalgamation (i.e. the 'joining') of distinct data files with centralised control. At the logical level there is no dependence on a particular storage structure or access method. This simplifies the applications code and makes it more readable. It facilitates development in that concentration is on the concepts rather than on matters such as byte positions.

Database systems typically have data dictionaries which allow access to the data through the high-level query languages which are usually provided.

The user can perform all functions of retrieval, creation, deletion and modification without knowing any details of the physical data storage.

There is no need to be concerned about formats, to pack or unpack records, to maintain link lists, to refer to specific volumes or for any of the specifics that accompany languages such as Fortran, Cobol, Basic or the like.

There can be a price to be paid for these advantages. Often it is slower execution. This need not be the case. (There is much confused thinking around that fails to distinguish logical and physical representations and which estimates operating speeds on the basis of logical structures -- a nonsense. It is unfortunately true that the databases as they are often set up do not take advantage of the storage structure and index options that will permit rapid response to be obtained.)

Note the difference between the requirements for an operational procedure and for a research exercise. In the former, consideration of efficiency and speed loom large. In the latter, it is the ability to work at the

appropriate intellectual level that is important. The reader with experience of procedural systems that require the user to concentrate on byte positions will recognise the inability to focus on the formulation of ill-defined problems that comes from this. In research such as that reported in this thesis, much of the computing effort was directed to developing strategies that were subsequently discarded or extensively developed as the understanding of the situation deepened. Flexibility and ability to work at the problem rather than program level were important. Dbms was most useful.

The reader should recognise that the results given here, and the algorithms provided, represent only a small part of the total computing carried out. If it had been clear that these results were what was useful, then the choice of coding procedure could have been considered in the light of that requirement. It may have been that a pattern matching programming language such as the UNIX awk would have been preferred to Ingres and Quel. However, that was not the situation faced when the Dbms road was followed.

The choice of the researcher was for a relational database and this is discussed further in 3.3.1 Data Analysis.

At this point in the discussion it is apt to note that the only Dbms available were relational. Had there been a

choice between types of database, the decision would have been made on the basis of the capability of the high level query language. It was important to achieve the isolation from physical considerations that is available with a mapping language such as Sequel (or SQL) or the relational calculus of Quel. The latter is preferable to the former on the grounds of superior writeability. Though SQL is virtually 'the' preferred high level query language these days, its strength is in its readability and not its writeability.

Analysis required the facility for data definition, retrieval, update, and integrity verification that is not bound and defined by the application software.

The data dictionaries found in most database management systems provides this flexibility. Data dictionaries are effectively a database in their own right. A dictionary contains "descriptions of other objects in the system rather than simple 'raw data'...A comprehensive dictionary will also include cross-reference information showing, for instance, which programs use which pieces of data, which departments require which reports."<sub>1</sub>

In general, this researcher and the library management are only interested in the information in the database and what operations could be performed on it. It is recognised that such analyses are carried out infrequently

and thus computing efficiency does not figure highly. A database management system and a high level query language fit the requirements here.

Some of the basic notions of relational database management systems are included in Appendix 3.1 to facilitate the reading of subsequent sections.

### 3.3 Aspects of Computer Processing

#### 3.3.1 Processor Requirements

The analyses discussed here require substantial computing resources. This section looks at the needs met during the course of the study. Comment is directed also at the implications of using a processor that is "too small" for the task. Initially the analysis of circulation transaction data were to be analysed using an IBM mainframe and Fortran programs that had been developed.

The initial analysis was to be based on use (for each Dewey section under review) of the following transaction information:

1. call number (in expanded form for processing convenience),
2. borrower's identification numbers,
3. date borrowed, date due, date returned.

Processing options that were available included extraction of distributions and statistical measures including:

- a. number of transactions (for the period under review),
- b. number of titles with cumulations based on activity level,
- c. 'per-copy' loan information on a title basis,
- d. detail of the time interval over which there was substantial loan activity.

The analysis was intended to span ten years. This did not present any storage problems on the IBM mainframe and analysis using Fortran programs and various standard utilities, such as sort and tape handling routines, were available.

During the period of the research it became necessary to transfer the work to a VAX 780 running UNIX.

Although this was a large installation by both research and library standards, it did not have the capacity to handle the large files without some segmentation. It was necessary to preprocess the files further than was desirable and to work subsequently with the summaries this produced. Further, although high level languages such as

Fortran were available, the execution times involved in their use were prohibitive. Accordingly the UNIX commands sort, awk\*, etc. were used for the preprocessing. In effect there was a utility-oriented analysis rather than a program-oriented analysis.

In retrospect, it is apparent that the limitations encountered in the latter stages of this research are representative of those that would apply in practical library situations. It is thus worth dwelling on their implications. As the research was commenced with use of an installation without these drawbacks this project has shown clearly the implications of the restrictions.

\* Awk is a UNIX command that can carry out pattern matching exercises. It can only execute with a single pass on the data. This can give rise to substantial limitations.

### 3.3.2 File Development

Analysis of the tapes revealed a number of problems that required attention. These included:

1. Incomplete years of statistics, and in some cases full years of statistics unavailable.
2. Overlap of statistical data into the following year.
3. Changes in processing over the years had produced a variety of record formats.
4. Practices on allocation of borrower number had varied over the period of the analysis.
5. The records for several of the years had bad blocks (there had been no practice of regular reading and rewriting of the tapes).

There are processing implications of several of the above. Point (2) requires sorting to remove duplicates. It is also desirable to run off an analysis of loans on a daily basis to verify that there are no gaps (e.g. a missing week) in the data. Point (3) requires a preprocessing program that can handle a variety of input formats and produce only the one format for the analysis routine. Point (4) needed to be considered when the loans were

being classified by type of borrower. The algorithm used varied with the year being analysed. Point (5), though avoidable if care is taken with tape storage procedures, is, unfortunately, to be expected with long term studies.

It is valuable if the tape units and tape handling software used have sophisticated error recovery routines. This was the case with the IBM installation used initially and its absence posed major problems in the VAX 780 installation. Since libraries are so dependent on tape files, this feature is quite important to them. It is also helpful if the processor is equipped with multiple tape units since the absence of such devices can render 'proper' tape maintenance difficult and thus make it more likely that it will not be carried out. (Multiple devices can also assist processing such as large sorts.) The VAX 780 had only one tape unit.

To handle some of these problems, a series of programs were written in 'C' which stripped all transactions having particular Dewey or class numbers. The intention was to produce a set of records with the same structure for each year, adding 'year' to those cases which did not have it as part of their format. (This programming was not done by the researcher.)

Some of the early years did not have the date borrowed -- a field of interest to this research. The fields required included the complete call number (with copy number), borrower identification, and date detail. Other fields, such as fine action and short author/title were not used.

For every year that could be handled, transaction records for the part of the collection of interest, were put into separate files which were subsequently merged. (The need for separate intermediate files came from the different processing requirements used for each year.)

'Sorts' of the files were necessary at this stage for two main reasons. Firstly, there were transaction records where material had been borrowed in one year but returned in another year. This resulted in duplicate entries (i.e. one entry each year). Secondly, for some years there were partial tapes and these overlapped creating duplicate records. These duplicates needed to be removed (in this case duplicate means records with the same call number, borrower number and date returned). Thirdly, the sort was used to bring together everything relating to one title by eliminating copy numbers and location prefix from the sorted files.

In summary these 'C' programs and sorts provided:

1. The removal of duplicate transaction entries and deletion of prefix and copy detail with an output file sorted by title.
2. The selection of classes as required with provision of an output format that was the same for each year. This gave the call number (in expanded format) and date detail.
3. The arrangement by borrower-type determined from the first two digits of the borrower identification number; and arrangement by year of borrowing also identified by the first two digits of the date borrowed.

The 'undergraduate' borrower group was split into 'new' undergraduates (students enrolled in the year under review and the previous year) and 'old' undergraduates (students enrolled prior to the current and previous year). This could be deduced from the borrower identification number.

The strength of 'awk' is pattern matching and this proved very useful in preparing the data for input to the Dbms.

The classification of the borrowers was part of the 'awk' processing. So too was the summarising of activity on a

per title/year/borrower class basis.

'Awk' was also used to process summaries of activity on a titles/transaction basis by Dewey class.

The type of detail obtained here was:

Call number	Number of	Number of
<u>Class</u>	<u>Titles</u>	<u>Transactions</u>
330.01	20	79, 80: 31, 30
330.015	2	79, 80: 0, 1
330.0151	38	79, 80: 81, 87
330.0973	2	79, 80: 1, 2

1979/80 Activity Summary on title/  
transaction basis by Dewey class

Table 3.1

In addition to the title/transaction display, a total summary count was provided as per:

Dewey Section 330

Total number of titles active 257

Total Transactions 508 (Year 1) 534 (Year 2)

These summaries provided a comparative analysis by class within each section of the Dewey under review. Areas of

little or excessive use became quickly apparent and acted as a basis for selection of classes for further analysis.

### 3.3.3 Ingres Processing

Once the files had been established, they were then downloaded for processing on the Ingres (Interactive Graphics and Retrieval System) relational database system. A 'copy' command allowed transfer of UNIX files to Ingres files.

The high level query language provided is Quel. This frees the data user from concern about data structures and storage. The user concentrates on his problem and not on detailed aspects of the way it is processed.

Ingres Version 7 (1981) was available on the University of NSW Electrical Engineering VAX 780 computer.

It was considered appropriate for circulation data analysis because:

1. It can handle large files.
2. It supports aggregate operations such as:

Count -- Count of occurrences

Countu -- Count of unique occurrences

Sum	--	Summation
Sumu	--	Summation of unique values
Avg	--	Average(sum/count)
Avgu	--	Unique average (sumu/countu)
Max	--	Maximum
Min	--	Minimum

and has 'group by' capability.

3. There is a macro facility. This provides the user with the ability to tailor Quel to the current problems.

Part of user friendliness desirable when one is trying to determine how to approach an analysis (rather than when one is trying to carry out the analysis) is achieved if involved command sequences can be turned into procedures. For example, if the problem being studied is to determine a relation between a threshold and the number of titles meeting a condition, it helps to issue a command like

number (thresh = 5)

The investigator is concerned with the relationship and his primary focus is not on the method of calculation. This can be particularly useful where 'number' involves selection of tuples, grouping, selection of groups and then the counting operation. Effectively it helps to parametise the operations and then merely to pass

parameters to a defined procedure. The macro feature simplifies this and thus promotes thinking at an appropriate intellectual level.

In short, the macro facility allows strings of text to be removed from the query stream and replaced with other text. This facility's advantage is that it can be used to shorten search statements used often and consequently making the search strategy less cumbersome and more user friendly.

#### 3.3.4 Circulation Relations

It was neither necessary nor efficient to look at individual loans when studying the usage patterns of most interest in this thesis. Accordingly, the processing was carried out on summaries.

There were two main studies made. One was concerned with usage on a per title basis over an extended period; the other looked at one year only. Naturally the amount of detail differed in the two cases.

With the longitudinal study the attributes required were total transaction counts for the individual years. With the shorter time period it was practical to study the user types. Naturally it was possible to link the two files and to permit a selection on the basis of one set of

relations to be joined to the other relations.

Specifically, the relations were:

1. RELATION FULLSTATS (Class, cutter, cterm, YR1, YR2, YR3, etc)

PRIMARY KEY: Class, cutter, cterm

ATTRIBUTE: CLASS -- the domain includes all class numbers in the areas chosen for analysis, i.e. 520.1, 330.9, etc.

ATTRIBUTE: cTERM -- includes book number

ATTRIBUTE: CUTTER -- includes cutter numbers where relevant

ATTRIBUTE: YR1, YR2, ETC -- includes all circulation transactions for the years under review (in the extended period analysis) for each title.

RELATION FULLSTATS -- Example

Attributes					
Class	cTerm	YR1	YR2	YR3	
62011	2	12	17	24	} Tuples
6201101	3	14	10	56	
62002425	1	15	2	10	

This relation was used in the extended period analysis.

2. The array established for the two-year study was a development that included attributes for use by type of borrower.

RELATION YEAR (class, cutter, cterm, NYR1, NYR2, nal,  
na2, np1, np2, nn1, nol, nn2, no2)

ATTRIBUTE: CLASS -- includes all call numbers in the  
(CHARACTER) chosen area for the two year  
analysis.

ATTRIBUTE: CUTTER -- cutter number where applicable.  
(CHARACTER)

ATTRIBUTE: CTERM -- book number

ATTRIBUTE: NYR1 -- represent the years being  
analysed and provide a total  
transaction NYR2 figure for each  
tuple under review

ATTRIBUTE: nal -- Academic use for year 1 and 2  
na2 under review (number of  
transactions)

ATTRIBUTE: np1 -- Postgraduate use for year 1 and 2  
np2 under review

ATTRIBUTE: nn1 -- New\* undergraduate use figure  
for year 1 under review.

\* (New is defined as current year 1st  
year enrolments and the previous  
year enrolments)

no1        -- Old\* undergraduate use figure  
            for year 1 under review

\* (Old is defined as undergraduate  
enrolments  $\geq$  2 years)

ATTRIBUTE: nn2        -- New undergraduate use figure for  
                      year 2 under review

no2        -- Old undergraduate use figure for  
            year 2 under review

This array was used in analysis which gave a comparative  
review of use by academics, postgraduates and  
undergraduates.

### 3.4 Conclusions

The sections in this chapter which deal with processor requirements, file development and Ingres processing provide only a summary of what occurred in development of files and their subsequent analysis.

It should be remembered that, at each stage of the file stripping, file development and file processing, a number of alternative methods were used and discarded as not meeting the requirements of the research. Some methodology in fact needed to be adjusted to be consistent with efficient processing.

The review, as it appears here, is the final version of the methods employed and was deemed to have been the most appropriate for this research's requirements.

### 3.5 Reference

1. Date, C.J. An Introduction to Database Systems,  
3rd ed. Reading Massachusetts, Addison-Wesley, 1981  
p.27.

## APPENDIX 3.1: RELATIONAL DATABASE STRUCTURE

### 3.1.1 Definition

"Given a collection of sets  $D_1, D_2, \dots, D_n$  (not necessarily distinct),  $R$  is a relation on those  $n$  sets if it is a set of ordered  $n$ -tuples  $(d_1, d_2, \dots, d_n)$  such that  $d_1$  belong to  $D_1$ ,  $d_2$  belongs to  $D_2, \dots, d_n$  belongs to  $D_n$ . Sets  $D_1, D_2, \dots, D_n$  are the domains of  $R$ . The value  $n$  is is the degree of  $R$ "<sub>1,2</sub>

Relational models typically build on the familiar rectangular array (or table). The terminology associated with the rational data bases follows the usage of set theory.

### 3.1.2 Relation

In common practice this is the table. A relation is defined over sets  $D_1, D_2, \dots, D_n$ . Each of these sets gives the domain of one attribute of the relation.

### 3.1.3 Domain

Is the pool of values from which actual values of one attribute (i.e. values is one column of the table) are drawn. Domains have within them all components that are valid.

Note, however, that all entries from a domain need not actually appear in a relation at any one time.

### 3.1.4 Attributes

Attributes have values that are drawn from the domains and appear as the columns in the rectangular array.

ATTRIBUTE: BOOK #, Domain Book Number

ATTRIBUTE: BOOK-Name, Domain Book Name

This description of the database is called a schema. The database has:

- 1 Relation (Books)
- 2 Attributes (Book number, book name)
- 2 Domains (Book number, book name)

### 3.1.5 Tuples

Tuples appear as rows in the rectangular arrays. A crucial feature of relational data structure is that associations between tuples are represented solely by data values in columns drawn from a common domain.

A relation is concerned with representation of information. If you are told the item with call number 330.1/11 had five loans in a period and you are told that 330.1/12 had ten loans in the same period, then you have been given some information. You do not have any more information or any less information if you are told that 330.1/12 had ten loans and 33.1/11 had five loans. There is no information in the order; there may be convenience in being given information in a certain order but no more information.

Relations and relational data bases are on about information. Given this, it is reasonable that a relation has the following properties (following Codd<sub>3</sub>):

1. Each row represents an n-tuple of R.
2. The ordering of rows is immaterial.
3. All rows are distinct.

4. The ordering of columns is significant -- it corresponds to the ordering of  $D_1, D_2, \dots, D_n$  of the domains on which  $R$  is defined.
5. The significance of each column is partially conveyed by labelling it with the name of the corresponding domain.

Since a relation is concerned with giving information, then these requirements can be seen to follow. For example, repeating a tuple gives no new information and thus repetition of tuples is not permitted. Likewise, there is no additional information in order. Consequently ordering is unimportant in relations. The column sort also adds no more information but it is necessary to know what the order is.

It is common (though not necessary) to require that each domain value is a single non-decomposable data item. This requirement relates to normalised relations in first normal form. Depending on the analysis to be done it may be necessary to have access to parts of the domain. For example, depending on research requirements, access to the call number of a book as a single data item may be sufficient. However, in some cases access to Cutter number and copy number would require that the domain be decomposable.

Often this amounts to the ability to specify string requirements. If string processing is not well supported by the Dbms, it may be simpler to decompose the data on input, i.e. in the above example to consider the call number as a concatenation of values for attributes like class, cutter, book number, copy number.

The crucial feature of the relational data structure is that association between attributes is represented solely by data values in columns drawn from a common tuple.

#### A Rectangular Array

Relation			Year			
Class	Cutter	Cterm	Nyr1	Nyr2	nal	na2
330942		121	3	14	1	0 ←
330973		107	2	15	0	0 ←
330973		76	2	15	0	1 ←
620		70	1	16	0	2 ←
6201		26	3	12	1	2 ←

} tuples

Table 3.2

Class -- class number

Cterm -- copy number

Nyr1 -- Transactions for Yr 1 for all users

Nyr2 -- Transactions for Yr 2 for all users

nal -- Transactions (academic) for Yr 1

na2 -- Transactions (academic) for Yr 2

## Keys

Tuples within a relation need to be uniquely identified. Within each relation there is at least one collection of attributes with values such that the combination is unique and consequently it can be used to identify tuples. For example, if circulation records are under scrutiny the call number of the item and the borrower number and the date of borrowing together identify the transaction. These three attributes thus provide a key for identifying tuples.

Since each tuple contains a distinct set of attribute values, each tuple can be distinguished from others within the relation.

For example, in relation of Table 3.2 (p. 145) the attributes 'class' and 'cterm' of the relation year can be used as the key. Each tuple has a distinct value for this attribute combination which will distinguish it from all other tuples in the relation. This combination is a way of distinguishing tuples (i.e. they have unique identification purposes) and it is thus a candidate key.

One candidate key is allocated as the key and is often known as the primary key. (It is necessary that the attributes of the primary key be fully defined.)

As shown, the primary key is not restricted to a single attribute and may be a combination of attributes that when taken together meet the unique identification property. Note that the combination of attributes must be the same for all tuples of the relation. Candidate keys that are not used as primary keys are referred to as alternate keys.

## APPENDIX 3.2: INGRES COMMANDS

### 3.2.1 Quel

'Quel' (QUErY Language) operates within Ingres and is a complete query language freeing the user from concern for how data structures are implemented and what algorithms are operating on the data.

Some of the major commands and features used in Quel are given below.

### 3.2.2 Copy

Copy allows the user to move data between Ingres and standard UNIX files. Direction is either "into" or "from". The format must be written for each domain and is a description of how it will appear (or appears) in the UNIX file. The relation must exist and have domain names identical to the ones appearing in the "copy" command. However, formats need not be the same in the domains and UNIX files as "copy" will automatically convert the data types.

It has two inconvenient features:

- i. "Copy" will cease operating at the first error,

- ii. When specifying the "filename", the entire UNIX directory path name must be provided.

This is because Ingres operates out of a different directory than the user's working directory.

The first of these may sometimes be a blessing in disguise; the latter is a nuisance.

### 3.2.3 Print

Copies a relation onto the user's terminal, formatting it as a report. It is a stylized version of "copy"; and no formatting is required on the part of the user.

One of its major drawbacks is its inability to print directly to a line printer. To obtain hard copy it is necessary to write to an output file and then to dispose it to a printer.

Other disadvantages are:

- \* Does not handle long lines of output correctly as there is no wrap around.
- \* Domain names will be truncated to fit into the specified width.

Its major advantage is that the system does the formatting, requiring no input from the user.

#### 3.2.4 Help

This provides information about how to use Ingres or about relations in the database.

It may be used to provide sections of the manual.

However, its major use is to give information about all relations in the database currently in use, or about specified relations.

It provides the following information:

Relation: CC

Owner: Crystal C

Tuple width: 44

Saved until: 12 Nov 1984

Number of tuples: 7

Storage structure: Compressed page heap

Relation type: User relation

<u>Attribute name</u>	<u>Type</u>	<u>Length</u>
Class	C	23
Term	C	11
n76	i	2
n77	i	2
n78	i	2
n79	i	2
n80	i	2

Once in Ingres, the user is reliant on the system defaults for precision and these may often be deemed excessive or inadequate, depending on how the system was set up. There may be advantages in changing these during the start up procedure.

### 3.2.5 Quel Syntax

Within 'Quel' there are syntactic classes from which the constituent parts of Quel statements are drawn. Some of the major ones used in this research are listed below.

#### i. Comparison Operators

=, >, <

#### ii. Logical Operators

not, and, or

### iii. Aggregate Expressions

Used to provide aggregations over sets of tuples.

The major aggregate operators are:

```
count -- count occurrences
sum   -- summation
avg   -- average (sum count)
max   -- maximum
min   -- minimum
```

Evaluation is to a single scalar value. Average and sum operate on numeric values; count, max, and min permit a character-type attribute as well as a numeric type. A small amount of care is required in deciding whether a variable such as `nyrl` is integer or character. In the latter case arithmetic operations are not possible.

#### 3.2.6 Range

A 'Quel' interaction includes at least one 'Range' statement of the form

```
RANGE OF [variable-list] IS [relation-name]
```

The symbols declared in the range statement are variables which will be used as arguments for tuples. These are called tuple variables. The purpose is to specify the relation over which each variable ranges.

Only ten variable declarations may be in effect at any one time.

### 3.2.7 Retrieve

Retrieve will get all tuples which satisfy the qualification and either display them on the terminal or store them in a new relation (if specified).

If all domains are to be retrieved, the keyword 'all' can be used.

Sample format is:

```
RETRIEVE [(into)relname] (target list)[where Quell]
```

e.g. retrieve into temp (e.all) where e. nyrl > 10

This must be preceded by a range statement, e.g. range of e. is year.

The search statements using retrieve syntax can often be long and cumbersome. Since they must appear for each search done on the data, re-input of the retrieve statement is necessary. If a number of comparisons are being made within the one retrieve statement, the syntax and input of the statement may become the primary concern of the users, forcing them to think at the wrong level,

i.e. at the syntax and input level rather than at the problem level.

Ingrès has developed within Quel a function known as 'Macros', which provides the user with the ability to handle retrieve statements easily.

The Macro facility provides the mechanism for tailoring 'Quel' to the user's own tastes. It provides the facility to remove strings of text from the query string and be replaced with other text. Also, some built-in Macros change the environment upon execution.

The Macros can be defined using the special command "define".

The usefulness of the Macro facility is the ability to shorten the retrieve statements and to replace attributes being tested easily.

Sample Macro statements are:

Macro 1

[define; rt; retrieve]

Macro 1 is a simple macro designed to assist in reducing the amount of typing required but has no intellectual content.

Macro 2

```
*      The frequency macro (as of 3/1/86)      *
*
*      Use this as:                             *
*      frequency relation attribute              *
*
[define; frequency $1 $2;
range of xyz is $1
destroy zza
destroy frq1
destroy frq2
destroy frq3
index on $1 is zza ($2)
retrieve into frq1 (value = xyz.$2, cases = count (xyz.$2
by xyz.$2))
range of zx is frq1
retrieve into frq2 (zx.value, zx.cases, trans = (zx.value
* zx. cases))
range of zy is frq2
range of zz is frq2
retrieve into frq3 (zy.value, zy.cases, zy.trans, ctrans
= sum (zz.trans), ctit = sum (zz.cases))
Print frq3]
```

Macro 2 is an operational macro. It has intellectual content and is designed to provide calculations, and to store and reformat data.

References: Appendix 3.1

1. DATE, C.J. An Introduction to Database Systems,  
3rd ed., Reading, Massachusetts, Addison-Wesley, 1981,  
p. 83.
2. CODD, E.F. A Relational Model of Data for Large  
Shared Data Banks. Communications of the ACM v.13 (6)  
June 1970, p. 379.
3. Ibid.

CHAPTER 4: THEORY DEVELOPMENT

	<u>Page</u>
4.1 Introduction.	163
4.2 'Core', its Definition and its Application.	164
4.2.1 Applications for 'Core'.	164
(i) Weeding of Collections.	164
(ii) Relegation.	165
(iii) Loan Periods.	167
4.2.2 'Core' -- A Transactions Activity Distribution Approach.	170
4.3 Short-term Analysis.	174
4.3.1 Distribution.	174
4.3.2 'Total' or Group Analysis	
-- Some Measures.	185
4.3.3 'Total' or Group Analysis	
-- Results.	196
4.3.4 Allowing for different Group Characteristics.	206
4.4 Analysis and Comparison of Results.	210
4.4.1 Results for the Varying Threshold Approach.	210
4.4.2 Results for the Group Characteristic Approach.	212
4.4.3 The Location Problem.	216

	<u>Page</u>
4.5 Long Term Analysis.	226
4.5.1 Distribution.	228
4.5.2 Maximum Transaction Level Analysis.	233
4.6 In-house Survey Results.	239
4.6.1 Survey Limitations.	239
4.6.2 Cluster Analysis.	240
(i) Type 1 -- Several copies used in one day.	243
(ii) Type 2 -- One copy only in use on a given day.	244
(iii) Type 3 -- One copy use over several days (with no date gaps).	244
(iv) Type 4 -- A title with three or more uses over seven days.	245
(v) Type 5 -- Several days use of a title with no more than a 3-day interval between uses.	246
(vi) Type 6 -- One copy use only with at least a 3-day interval between uses.	246

	<u>Page</u>
4.6.3 General Conditions in Clusters.	247
4.6.4 Results of In-house Survey.	247
(i) Conjectures.	248
4.7 Open Reserve Room Results	258
4.8 References.	263

		<u>Page</u>
Table 4.1	Sample cumulative distribution	174
4.2	Activity level on title/transaction and discipline basis for titles used once 1979	177
4.3	Activity level on title/transaction and class basis for titles used once 1979	178
4.4	Activity level on title/transaction and class basis for titles used once 1980	179
4.5	Cumulative transaction thresholds 1979	197
4.6	Cumulative transaction thresholds 1980	198
4.7	Cumulative transaction thresholds on a class basis for Economics/Management 1979	199
4.8	Active titles on a discipline basis 1979/1980	202
4.9	Efficiency, relevance, overlap on a discipline basis 1979/1980	204
4.10	Class efficiency, relevance, overlap for Economics/Management 1979/1980	205
4.11	Efficiency, relevance, overlap (Kent-type) on a discipline basis 1979/1980	209
4.12	Overlap and Ratio 1 measures for disciplines and approaches basis 1979	222

	<u>Page</u>
4.13 Overlap and Ratio 1 measures for class and approaches basis	223
4.14 Sample cumulative distribution for maximum use for all years reviewed for all undergraduate use	229
4.15 Cumulative transaction thresholds for maximum use for all years reviewed for all undergraduate use	231
4.16 Title/year distribution for all undergraduate loans with maximum transaction level $\geq 18$ for all years reviewed	233
4.17 Sample in-house patterns representing cluster types 1-6	242
4.18 a - d Class analysis of title/occurrences for each type of cluster	251 - 255
4.19 In-house survey results by class for titles, transactions and one use only titles	256
4.20 In-house survey results for one use only titles as a percentage of total titles and total transactions	257
4.21 Reserve Room class results	259

## Figures

	<u>Page</u>
4.1 Overlap of titles (old and new core)	188
4.2 Notion of relevance	193
4.3 Notion of efficiency	195
4.4 Overlap (old, new, total core)	218
4.5 Overlap (combined core)	224
4.6 Scatter diagram for $R_1$ and $R_2$	236

## Equations

4.1 Overlap (old/new)	188
4.2 Relevance	191
4.3 Efficiency	194
4.4 Overlap (old/new/total)	218
4.5 Overlap (combined core)	224

## Graphs

4.1 -	268 -
4.7 Frequency distributions	274

## Appendices

4.1 Undergraduate distribution for 1979	275
4.2 Terms used in the procedure	279

#### 4.1 Introduction

This chapter presents results of both short-term (one or two periods) and long-term (several periods) analysis of circulation records. The discussion is largely related to determining 'core'.

The first section (4.2) surveys situations where the notion of 'core' is used and addresses the characteristics that it should have. Sections (4.3-4) and (4.5) present the short-term and long-term results respectively. In large part these discussions look at the acceptability of approximations to the analysis of behaviour of individual user groups. It is recognised that the detail available to this thesis will seldom be found in practice.

Accordingly the extent to which the notion of 'core' can be related to the precision of group classification and the sensitivity of the results are important issues.

A discussion of the conclusions is provided in Section 5.2 where an assessment is made of the adequacy of analyses that are more simply based than those provided here. It effectively addresses the question of whether one needs more than would normally be provided. Sections (4.6) and (4.7) deal with the results from the surveys of in-house and reserve book collections' use. These complement the circulation discussion of 4.3-5.

## 4.2 'Core', its Definition and its Application

Research over the years has considered whether statistical procedures can predict accurately which books will be used. This has significant implications for acquisitions but it is also relevant once the item is included in the collection.

The analysis of use patterns presented in this thesis, follows the tradition and it can be applied to weeding, relegation, the setting of loan periods and to location of items within the collection.

### 4.2.1 Applications for 'Core'

#### (i) Weeding of collections.

The approach to weeding depends on general requirements of the patrons, the type of collection, the problems the library faces and the 'return' required from the weeding exercise.

Slote<sub>1</sub> mentions that weeding increases circulation, and the speed of access improves accuracy in retrieval (by minimizing irrelevant material).

A variety of weeding procedures have been devised; these include subjective weeding, and algorithmic techniques

based on age of the item, date of publication and shelf-time period (the time the book remains unused on the shelf). Many papers have used mathematical forms involving the shape of distributions, etc. Combinations of these approaches have also been used.

An alternative approach is based on the need of libraries to maintain at least minimum recommended standards (i.e. basic collections) and balanced collections.

This variety in approaches has resulted in recommendations varying from retaining all works currently in the collection, to maintaining the collections at predetermined physical sizes. Recently there has been emphasis on maintaining stocks that maximize use rather than variety of treatment of a topic. Space limitations and budgetary pressures have been important in determining the rate of growth of collections. All of these considerations have been incorporated in the formulation of collection development policies and thus weeding policies.

(ii) Relegation.

The transfer of material from primary storage (open stacks) to secondary storage (closed stacks) has been the subject of much debate.

A key issue is the impact on service of removal to secondary storage. Grieder<sub>2</sub> was one of the first to consider the value of being able to estimate this. He did so by using the date of last circulation. Grieder's research at the Stanford University Library found that 39.2 per cent of the books in the study had not circulated in the previous 15 years. The same circulation performance would have resulted if these works had been located in secondary storage. In a two-week study of books circulating he found that some 4 per cent and 3 per cent of the requests during those two weeks would have come from storage.<sub>3</sub>

The relegation problem has been expressed in terms of cost of storage and the optimal sizes of libraries by Buckland.<sub>4</sub> He extended the comparison further by examining storage and inter-library loan costs.<sub>5</sub>

The purpose was to determine what combination of acquisition and discarding policies would minimize the sum of the costs of acquiring titles, storing titles until discarded, inter-library loan costs for items discarded and inter-library loan costs for items never acquired in order to achieve the optimal combination of speed of service and cost of service.

It is evident that knowledge of circulation statistics is a necessary precondition to the carrying out of such costings.

This thesis looks at a relegation type of problem, the choice between research and undergraduate locations for an item. The basic issue here is similar to the open/closed stack case.

Inappropriate location can substantially reduce circulation and it will render in-house use impossible and thus diminish the usefulness of an item.

#### (iii) Loan Periods

Use patterns give insight when choosing the loan periods. Note that there are a number of classification approaches that can be relevant; all can use circulation information though what is relevant is not always the same.

The two basic strategies are to have an item-dependent or borrower-dependent loan period.

With the former, the determinant can be the item location. For example, there may be different loan periods for items located in different sub-libraries. Here the problem devolves to one of deciding location for the item and its circulation history is important.

Alternatively, different categories of borrower may have different privileges. In determining or reviewing these it can help to know the history of renewals, duration of loans, etc. Thus, once again, circulation information is relevant though now it is the detail for borrower classes. Here too, we have the problem of knowing the role that the borrower was filling when the loan occurred. For example, an academic may borrow items as a parent, as a person interested in a particular sport, as a tutor, or as a researcher. It would be appropriate to link the loan period to the role but this is not possible. The use of the patron's position classification needs to be recognised as a surrogate for the role for a particular loan. The mapping between the two will be fuzzy and, once again, there is a need for interpretation of any position-related statistical analysis.

There can, of course, be various combinations of these approaches. The statistical data must then be appropriate to the problem. For example, if it is conjectured that the loan period will be a function of both the borrower type and the item location, then details of borrowings of user groups from the various sub-libraries are needed.

The discussion of this thesis can be applied to deciding the location for an item. Accordingly, it is relevant to one of the main types of loan period problems.

The aim of some variable loan period research has been to allow users to borrow items for long periods if they are not in demand. Thus, 'serious' students will not be inconvenienced. The Lancaster Study (Buckland)<sub>6</sub> used the categorisation of 'popular', 'very popular' and 'other books' in determining loan period categories. It is worth recalling, however, that Lancaster needed to review a policy where material was available for loan over a three-month period. Accordingly, there was scope for considerable gain from the exercise.

In cases where loans are available for one week, or where recall of material is possible if on loan for longer periods, the variable loan period exercise is not so relevant and does not offer such significant gains in the use of material.

It is worth recalling the characteristics of loans to undergraduates in a university library and to note their differences from those of other users in other libraries. Although some material will always be 'very popular', it is difficult to deal with high moving material that is essay-related. This material can display high activity pulses in some periods and be relatively little used at other times. Thus, although these items are candidates for short loan periods and copies need to be held in reserve collections during the peaks, they may not be easily recognised and classified before the high demand

pulses are encountered. The items may only be known after the peak has occurred. This is particularly important if the duration of the pulse is only a matter of a few weeks. The detection of such high movers can be assisted by analysis of circulation statistics from previous years since use patterns are often highly correlated. In a sense, this is a type of variable loan period problem but it is one that has figured little in the literature.

#### 4.2.2 'Core' - A Transaction Activity Distribution Approach

A title's popularity will change in the course of its life. When the loans drop significantly for a title in an undergraduate collection, then it could be appropriate to move the item to a research collection and, possibly, make it borrowable for longer periods. At the time of such relocation, there is need to consider the number of copies that should be kept. Notions of weeding, relegation/relocation and the setting of loan periods thus are inter-related. The selection of titles for such examination brings up the matter of 'core'.

In one definition 'core' is said to be:

"books that would satisfy 95-99 per cent of the present demands made upon the entire present collection."<sup>7</sup>

Other definitions are similar in concept and relate 'core' to diminishing returns theory and the shape of the distribution. For example, it can be that 20 per cent of titles satisfy some 80 per cent of transactions. On a more mathematical basis, 'core' can be defined as relating to the region of linearity in the S-shaped use distribution curve.<sup>8</sup>

These definitions are variations on a theme and are refined here as specifying 'core' in terms of a transaction distribution curve.

The view taken in this thesis does not affirm that one definition gives 'the' correct approach. Rather, it is argued that 'core' is a concept that is useful and the analyses undertaken should not be determined on the basis of an 'arbitrary' figure such as the 95 per cent, but should be based on the problems that are to be solved.

For example, in dealing with undergraduate collections, a major concern can be bulk use and the supply of monographs to support reading lists. 'Warehouse-type' procedures may then be useful. Here it is appropriate to consider 'core' in relation to a percentage of total transactions since the specific follow-up questions involve the adequacy of holdings of a comparatively small number of titles. As suggested by Slote,<sup>9</sup> with a percentage approach the figure

used depends on the library's objectives and the possible unfavourable impact of impeded use of certain titles on the clients.

In a library which has a variety of clientele, e.g. a university library servicing early undergraduates (bulk borrowers), later year undergraduates, academics and post-graduates (limited specific use), it may be necessary to look separately at the requirements of each group.

A concept that has not been explored in detail is whether the behaviour of early and late undergraduates is different. If it is, then it may be important to determine who is responsible for the loans. This research examines whether it is necessary to consider the particular user group when analysing data, or whether the 'total' use statistics are sufficient if the focus is on the undergraduate group.

The analysis considered undergraduates in two disjoint categories, viz:

- 'old' undergraduates. This included students who had been enrolled two years or more prior to the year of use under review.
- 'new' undergraduates. Current year plus the previous year enrolments to the year of use under review.<sup>10</sup>

In addition, some analyses looked at 'total' undergraduates; that is, it included all undergraduate students.

There was also an 'academic and post-graduate' category of user. The loans of this last group were excluded from the analysis reported here to permit focusing on the undergraduate group.

It was considered appropriate to determine whether each user group required a different 'core'.

### 4.3 Short Term Analysis

This section presents the results from analysing one year's data. Section 4.4 gives a complementary discussion that looks at the behaviour over several years.

#### 4.3.1 Distribution

The initial process in establishing 'core' on a distribution basis is to generate the cumulative distribution for activity in the area by the users of interest. One such is:

<u># Trans</u>	<u># of Titles</u>	<u>Cumulative</u> <u>Titles</u>	<u>Cumulative</u> <u>Trans</u>
0 <sub>11</sub>	922	922	0
1	1115	2037	1115
2	465	2502	2045
⋮	⋮	⋮	⋮
44	2	3340	8202
57	1	3345	8455
⋮	⋮	⋮	⋮
114	1	3352	8863

Table 4.1 Sample Cumulative Distribution for Title/  
Transaction Activity for all undergraduates  
for Dewey Sections 657/658 for 1980.

For the discipline as a whole, this analysis revealed typical diminishing return behaviour, e.g. a high number of titles with one transaction only and a few titles with high use levels.

Frequency distributions, such as Table 4.1, provided breakdowns for both titles and transaction use. It was possible to look at this activity in terms of total undergraduate use, or to concentrate on particular undergraduate groups. Titles borrowed by only one of these categories could be isolated. The method of determining activity based on titles and transactions for total and particular undergraduate use is explained in the following examples.

1. Percentage of total active titles represented by titles used once for 'total' undergraduates 1979\*

Total titles	9126
--------------	------

Titles not used	
-----------------	--

(zero transactions)	<u>2018</u>
---------------------	-------------

TOTAL TITLES USED	7108
-------------------	------

TITLES USED ONCE	3100
------------------	------

Percentage of total active

represented by titles used	=	$\frac{3100}{7108} \times 100$
once		

= 43.06%

2. Percentage of total transactions for a user group represented by titles used once for 'total' undergraduates 1979:

Titles used once	3100	
Total Transactions	31100	
Percentage	$\frac{3100}{31100}$	= 10%

- \* Figures are from the UG79 distribution of transactions/titles Appendix 4.1

The results for 1979 for titles with only one transaction:

	<u>Economics and</u> <u>Management</u>	<u>Engineering</u>
Total U/G	43.1%	42.8%
'Old' U/G	52.3%	49.1%
'New' U/G	52.1%	55.3%

Percentage of total active titles represented by  
titles used once.

Table 4.2a

	<u>Economics and</u> <u>Management</u>	<u>Engineering</u>
Total U/G	10.0%	10.9%
'Old' U/G	18.3%	16.2%
'New' U/G	13.5%	20.4%

Percentage of total transactions for a user group  
represented by titles used once.

Table 4.2b

Table 4.2 Activity level on a title/transaction and  
discipline basis for titles used only once in  
1979.

<u>Economics/Management</u>		<u>Titles 1979</u>	
	<u>'Total' U/G</u>	<u>'Old' U/G</u>	<u>'New' U/G</u>
Class 330.0	40.2%	50.2%	41.4%
657-658	47.7%	53.4%	57.8%
330.1	36.6%	51.0%	46.9%
330.9	26.3%	40.0%	35.0%
338	46.0%	58.0%	51.0%

Percentage of total active titles represented by titles  
used once (1979)

Table 4.3a

<u>Economics/Management</u>		<u>Transactions 1979</u>	
	<u>'Total' U/G</u>	<u>'Old' U/G</u>	<u>'New' U/G</u>
Class 330.0	10.0%	15.3%	17.0%
657-658	13.8%	19.4%	18.5%
330.1	7.3%	15.6%	11.8%
330.9	3.1%	9.6%	5.5%
338	11.5%	23.0%	14.6%

Percentage of total transactions for a user group  
represented by titles used once (1979)

Table 4.3b

Table 4.3 Activity level on a titles/transaction  
and class basis for titles used once in 1979.

<u>Economics/Management</u>		<u>Titles 1980</u>	
	<u>'Total' U/G</u>	<u>'Old' U/G</u>	<u>'New' U/G</u>
Class 330.0	38.0%	49.0%	39.6%
330.1	35.8%	47.6%	47.5%
330.9	26.7%	40.3%	32.8%
338	40.8%	55.9%	48.0%
657-658	45.8%	53.4%	51.8%

Table 4.4a

<u>Economics/Management</u>		<u>Transactions 1980</u>	
	<u>'Total' U/G</u>	<u>'Old' U/G</u>	<u>'New' U/G</u>
Class 330.0	5.0%	15.0%	5.0%
330.1	7.1%	15.25%	10.6%
330.9	3.4%	10.9%	5.2%
338	10.3%	24.38%	13.6%
657-658	12.2%	19.0%	16.4%

Table 4.4b

Table 4.4 Activity level on a title/transaction and class basis for titles used once in 1980.

If 'core' is defined on a transaction percentage level to include 95 per cent of transactions, using either total undergraduate loans or individual user group results (as in Table 4.2b), all titles borrowed in the year would need to be included in 'core'. All entries in Table 4.2b are above 5 per cent (i.e. 100-95 per cent) and the minimum transaction level that would be required before some active titles were not included in 'core' would be 90 per cent. This would apply for the 'total' undergraduate user group in the economics/management area. (The figure 90 per cent comes from 100-10 per cent, the latter being the figure in Table 4.2b.) With the exception of two cases (330.9 for 'total' undergraduates in 1980 and 1979), this also applies for the class breakdown in the economics/management area as shown in Tables 4.3b and 4.4b. The minimum transaction level that would be required before some active titles were not included is 97 per cent for both 1979 and 1980. Effectively this means that all items borrowed in a year would be included in this definition of 'core' determined from that year's loans. This may present a problem.

In some applications, for example, item relocation, 'core' may not be a useful concept if it includes everything that is active in the collection.<sup>12</sup> Note, however, that putting all borrowed titles into 'core' is not always a problem. If, for example, the exercise of interest is one of weeding, then the results of Table 4.2 lead to the

study of the multiple copy holdings of inactive titles. Further, the values in the table indicate that the decision on what constitutes 'core' is not sensitive to variations in the threshold from, say, 95 to 90 per cent.

There are, however, applications where 'core' needs to be only a portion of the active titles. One such is the review of the adequacy of multiple copy holdings of heavily used materials. Here there will be a small number of titles but the portion of the circulation that they represent will be significant. Here, the figure of 95 per cent is not appropriate.

To get a feel of how much the 95 per cent needs to be weakened, a spread of transaction levels (ranging from 80 - 20) was examined. (Note that the different percentages are not really alternatives. Rather, they represent situations where different values are appropriate and provide a range from 100 - 0 per cent.)

These transaction levels are determined from the cumulative distributions and illustrated in Table 4.1 and graphs 4.1 - 4.7. For each percentage cumulative transaction level, the corresponding transaction activity was determined and used in the algorithms to analyse use.

Apart from the setting of cumulative percentages to determine 'core', there were two other problems that

needed attention, viz, deciding how to handle the very high movers and questions of precision in the use of the distributions.

In the cumulative distributions for transactions there were a number of cases of titles that had more than 100 loans in a year. The percentage use curves (Graphs 4.1 - 4.7) and the percentage levels of transactions used for establishing 'core' have included these titles. In the economics/management area in 1979 there were 45 titles with more than 100 transactions and they accounted for 8 - 10 per cent of loans for the total user group. In 1980 the number of titles was 27.

Titles with this high level of circulation were considered to form a unique category, different from the remainder of the collection. There was concern that the inclusion of such titles, although only few in number, would distort the use curves and the cumulative transaction levels. The crucial point here concerns the aim of the exercise. If the library problem requires a global look that includes every title, then there is no case for a special category. When, however, the aim is to look at issues like weeding, location, loan periods and the like, then the very high activity group may obscure the examination of other titles. These titles are clearly not candidates for weeding or relocation. Their exclusion will not weaken the validity of such an analysis. Note that it is not

being said that there is always only one answer on the question of exclusion or not. These titles should be included or excluded as appropriate to the study in question.

A global view was required in this analysis, hence titles such as Samuelson's Economics, which were responsible for more than 600 transactions for the year were included. It should be remarked that there were no comparable titles in the engineering discipline. (The highest transaction level occurred with two titles and was of the order of 100 transactions/year.)

If weeding or relocation were the primary concern there would have been no problem in excluding the titles in the analysis.

It needs to be recognised that distortion might occur. The decision as to whether to include these titles or not is essentially a matter of what the library is currently examining and the development goals for that part of the collection under review.

The second point to note about the cumulative distributions concerns accuracy. Cut-off points on the distributions are not 'exact' because the distribution is only defined at integral values of transaction levels. (The smooth curves commonly shown are illustrative rather

than rigorous). The chances of the precise value of transaction activity corresponding to a point on the distribution is slight.

The important point is that this approach is intended to provide an initial selection of items that are relevant. It is intended that there will be a follow up examination and, thus, it is not necessary to have high precision in this initial operation. The 'error' in precision in most cases is between 1 - 5 per cent. This raises an important point of philosophy about the role of algorithms and procedures such as those advanced in this thesis.

Although there is a view such as that of Slote<sub>13</sub> which is critical of collections being weeded by professionals using judgment not rules, the need for a subjective element cannot be rejected.

THE APPROACH OF THIS RESEARCH IS TO PROVIDE MACHINE ASSISTANCE FOR SELECTING RELEVANT ITEMS THAT CAN BE EXAMINED. BY DOING THIS WE AVOID SOME OF THE LABOUR COST AND IMPOSE A CONSISTENCY IN THE TASK BEYOND THAT POSSIBLE WITH THE SUBJECTIVE APPROACH ONLY.

#### 4.3.2 'Total' or Group Analysis -- Some Measures

The approach normally followed for analysing undergraduate use data is to look at total behaviour. This 'total' approach goes unquestioned. All users are considered to be members of this 'total' group. In order to check whether this approach is valid, 'total' user figures were examined and compared with use figures for various specific user groups.

If users are to be considered in various categories, there are two important considerations:

- (1) When capturing data it is important to identify the user's group. It is not sufficient to collect only total loan data on a title basis.
- (2) The method of selection of titles for analysis is important. The major question here is whether titles selected on a percentage transaction cut-off basis for individual groups will be the same as those selected for all users with a comparable transaction percentage.

It is this second consideration that will concern us. To resolve it, a number of algorithms were designed to look at overlap in the use of titles. These identified titles used by particular user groups and established whether the same titles were of interest to several groups.

Activity in the economics/management discipline was analysed in terms of both the discipline as a whole and for individual classes. In the case of engineering, analysis was to illustrate rather than to examine in great detail what was occurring. In this latter case the concentration was at a discipline level.

The research questions that needed to be resolved were:

1. Are the titles of interest to 'old' and 'new' undergraduates essentially the same?
2. Are the titles retrieved by using 'total' undergraduate use the same as those resulting from the union of the 'core' for the two subgroups?

This second question can be re-expressed as two subquestions, viz:

- 2a. Are the titles that are in core for the 'total' undergraduate use substantially drawn from the combination of 'cores' for the user group?
- 2b. Are titles that are in 'core' for the 'total' undergraduate group a substantial percentage of the 'core' for the two subgroups?

To deal with Question (1), it was necessary to determine those titles in the 'core' for the two groups separately and then to compare the sets. One simple method of comparison is to compare the number of titles that are common to the 'core' for each group with the number of titles in the union of the two sets.<sup>14</sup>

Let

TOLD = number of titles in the 'core' for the 'old'  
undergraduate group

TNEW = number of titles in the 'core' for the  
'new' undergraduate group.

and

TTCM = the number of titles that appear in 'core'  
for both the 'old' and the 'new' groups.

(NOTE: The procedure given in Appendix 4.2 defines all  
terms used in equations 4.1 to 4.3)

The first measure of commonality (overlap) can be expressed as:

$$\text{OVERLAP (OLD/NEW)} = \frac{\text{number of titles common to both groups}}{\text{total number of titles in both 'cores'}}$$

$$\text{and} = \frac{\text{pop. of intersection of the two sets}}{\text{pop. union of the two sets}}$$

$$= \frac{\text{TTCM1}}{\text{TOLD} + \text{TNEW} - \text{TTCM1}} \quad (4.1)$$

The derivation of the expression (4.1) can be seen from the Venn diagram of Figure 4.1.

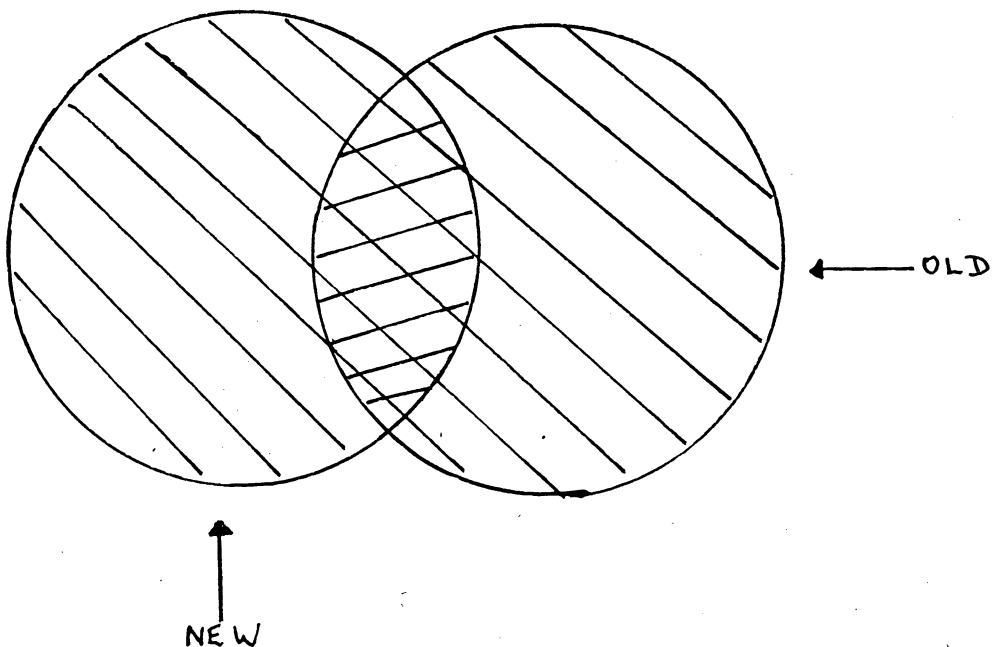


Figure 4.1. Depiction of overlap between titles in 'new' core and 'old' core.

In Figure 4.1, the area hatched (\\\\\\) represents the union of the core for the 'new' and 'old' undergraduate groups. The part hatched (////) represents titles common to these two cores.

The number of titles in the 'combined' core, (i.e. the number of titles in the union of the sets 'new' core and 'old' core) can be obtained from the number in each core separately. Note that the number of titles in the 'combined core' is not  $(TOLD + TNEW)$ . That expression double counts titles common to both cores. The number in the combined core is  $(TOLD + TNEW - TTCM1)$  where  $TTCM1$  is the area (////).

Naturally, when using a measure such as overlap (old/new) the cores must be determined consistently. This need not necessarily mean that the same percentage threshold applies for both groups.

The thresholds provide a way of grouping titles in increasing (or decreasing) order of transaction activity. The determinant of core need not be based on a threshold criterion though a threshold approach may yield the same groupings. In such a case, different thresholds or different percentage use figures may apply for consistent groupings.

It is worth noting the maximum value that may be taken by the overlap of Equation 4.1. This occurs when one set of titles is contained within the other set, then we have:

$$\text{OVERLAP} = \frac{\text{MIN (TOLD, TNEW)}}{\text{MAX (TOLD, TNEW)}}$$

Note that the maximum value for this overlap need not be 1 and it will only be 1 when the two sets are identical.

Question 2 (Page 186) involves an alternative to determining the core for the two groups individually, viz, evaluating it for the undergraduate group as a whole.

On first encounter this may appear to be fairly straightforward but the comment about the merits of different use percentages for the two subgroups needs to be noted.

When the undergraduate population in total is considered, the ability to allow for differences between subgroups is lost. If the answer to Question 2 is "yes", then we do not lose by following the simpler approach. If it is "no", then we will need to look at the difference between the sets involved. This situation can be appreciated best by looking at the subquestions (2a) and (2b).

One measure that can indicate the validity of the 'total' approach looks at the overlap between the core for 'total' and the cores for the two groups. If there is much in common between the sets of titles, then the 'total' core can be considered relevant to the union of the subgroups.

Thus we can define a "relevance" measure as:

$$\begin{aligned} \text{RELEVANCE} &= \frac{\text{Number of titles common to 'total' and 'combined' cores}}{\text{Number of titles in 'total' core}} \\ &= \frac{\text{TTCM} - \text{TTCM2}}{\text{TOTAL}} \quad (4.2) \end{aligned}$$

where:

$$\begin{aligned} \text{TTCM} &= \left\{ \begin{array}{l} \text{Number of titles in 'total' core and} \\ \text{'old' core} \end{array} \right\} \\ &+ \\ &\left\{ \begin{array}{l} \text{Number of titles in 'total' core and} \\ \text{'new' core} \end{array} \right\} \end{aligned}$$

$$\begin{aligned} \text{TTCM2} &= \text{The number of titles that are} \\ &\text{common to the 'new', 'old' and} \\ &\text{'total' groups.} \end{aligned}$$

TTCM - TTCM2 =       The number of titles that are common to  
                              'total' and to the combined core for  
                              the 'new' and 'old' group.

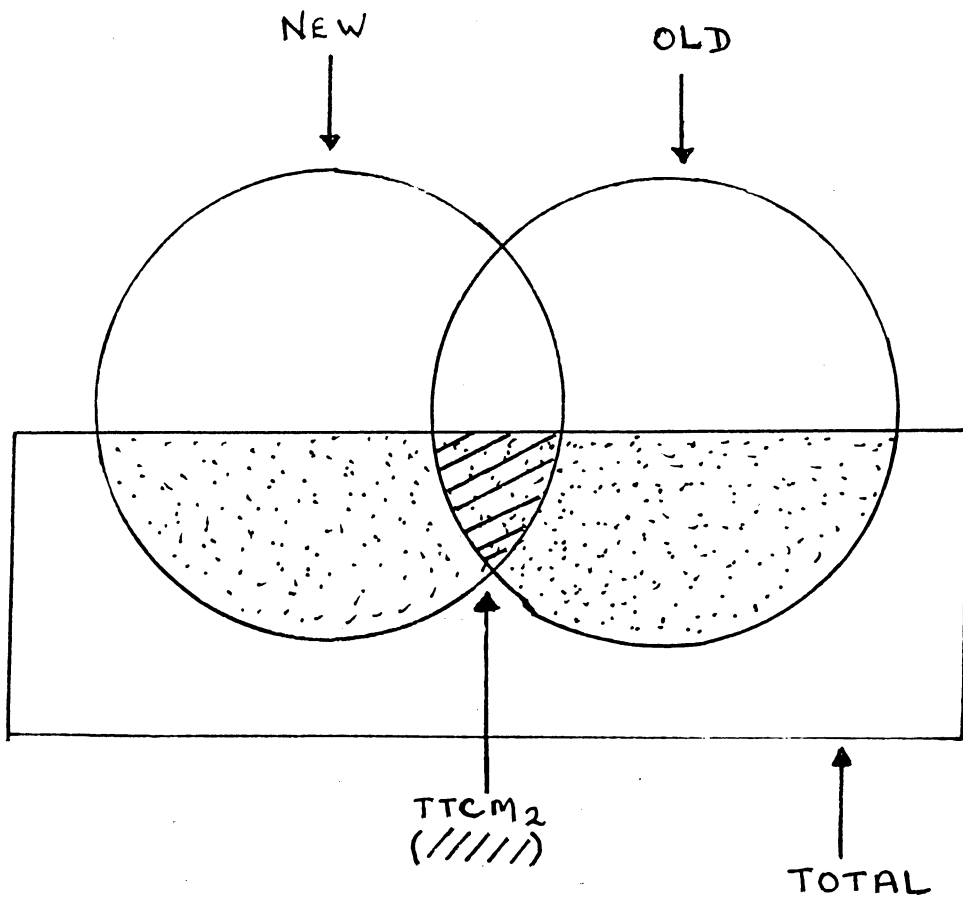
TOTAL                =       The number of titles  
                              in the core determined for the  
                              undergraduate group as a whole.

(Note the difference between TTCM and TTCM2. TTCM2 is the  
intersection of TTCM1 and TOTAL.)

Figure 4.2 gives a Venn diagram for the 'new', 'old' and  
'total' situation. The area shown dotted corresponds to  
the titles in 'total' that are also in the combination of  
the two group cores. The 'relevance' of Equation 4.2 is  
the proportion of the area of TOTAL represented by the  
dotted section, i.e.,

RELEVANCE \* TOTAL

= number of titles in 'total' core  
that are included in the 'combined' core.



[ (.....) is  $TTCM - TTCM_2$  ]

Figure 4.2 The Notion of Relevance

Question (2b) can be answered by looking at the intersection of the 'combined' core and 'total' as a percentage of the 'combined' core. If there is much in common between the two sets of titles in 'total' and 'combined' core, then 'total' can be considered an efficient predictor of the combined core.

Thus we can define an "efficiency" measure as per:

$$\begin{aligned} \text{EFFICIENCY} &= \frac{\text{Number of titles common to 'total' and 'combined' cores}}{\text{Number of titles in 'combined' core}} \\ &= \frac{\text{TTCM} - \text{TTCM2}}{\text{TOLD} + \text{TNEW} - \text{TTCM}} \quad (4.3) \end{aligned}$$

In the Venn diagram of Figure 4.3 efficiency is the ratio of the dotted to the hatched area.

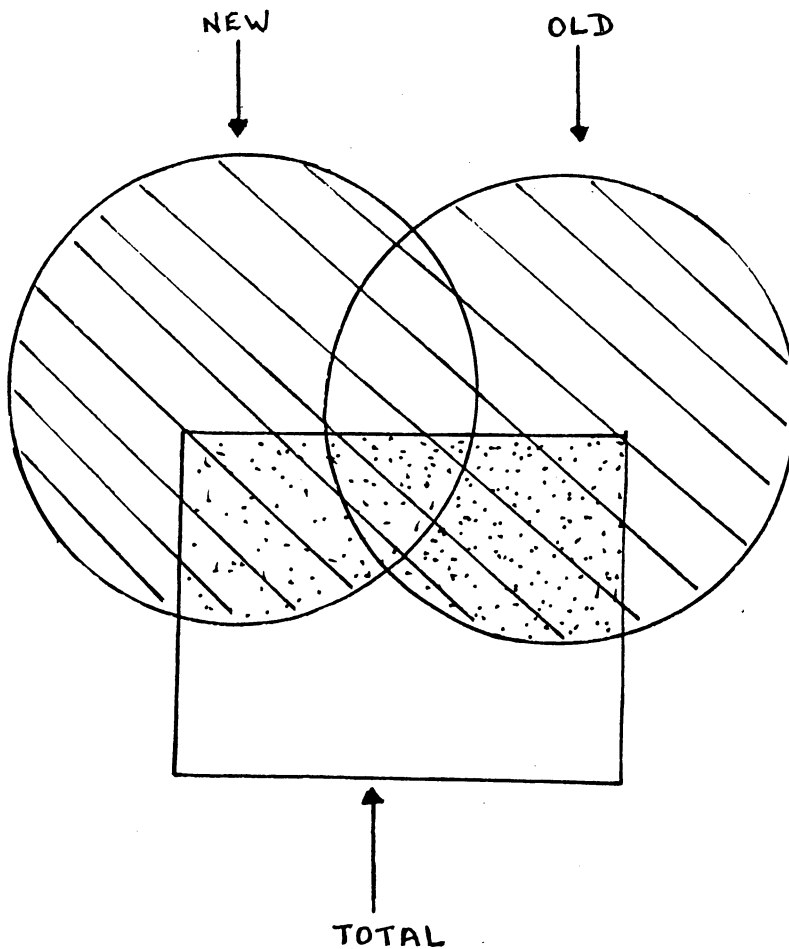


Figure 4.3 The Notion of Efficiency

#### 4.3.3 'Total' or Group Analysis -- Results

Equation 4.1 measures overlap between the 'old' and 'new' core on an individual group basis.

There can be different thresholds set for each group. It would often be the case (though it would not be mandatory) that roughly the same cumulative percentage use figure could apply for each group.

For each discipline and class reviewed these figures were established for cut-off points spanning the range of between 80 per cent and 20 per cent of cumulative transactions. This provided a detailed view across a broad range of activity and thus included values appropriate for many problems that may be encountered.

Note the point made previously that it is not proposed that figures such as 20 per cent and 40 per cent are alternatives to the 80 per cent cases; rather they are relevant in different situations.

From the cumulative transaction tables, various transaction thresholds were established for analysis of 'old', 'new' and 'total' undergraduate use. The discipline results were:

Cumulative %	'Old' U/G	'New' U/G	'Total' U/G	All
Transactions	Use	Use	Use	Users
80	1	1	2	2
60	2	4	5	9
40	6	13	14	30
20	15	42	45	45

Total Economics/Management Undergraduate Use.

Table 4.5a

Cumulative %	'Old' U/G	'New' U/G	'Total' U/G	All
Transactions	Use	Use	Use	Users
80	1	0 *	1	1
60	3	2	4	4
40	6	5	9	8
20	16	17	22	20

(\* The threshold '0' corresponds to 100 per cent. This was the closest that the 80 per cent could be approached.)

Total Undergraduate Engineering Use.

Table 4.5b

Table 4.5 Cumulative transaction threshold results for each user group Economics/Management and Engineering disciplines for 1979.

Cumulative % Transactions	'Old' U/G Use	'New' U/G Use	'Total' U/G Use	All Users
80	1	1	2	2
60	2	4	5	5
40	5	10	12	12
20	12	30	34	32

Total Economics/Management Undergraduate Use.

Table 4.6a

Cumulative % Transactions	'Old' U/G Use	'New' U/G Use	'Total' U/G Use	All Users
80	1	1	1	2
60	3	3	4	4
40	6	7	9	9
20	14	22	24	24

Total Engineering Undergraduate Use.

Table 4.6b

Table 4.6 Cumulative transaction threshold results for each user group Economics/Management and Engineering disciplines for 1980.

Thresholds for classes in Economics/Management 1979 were:

Cumulative %		'Old' U/G	'New' U/G	'Total' U/G	All
Transactions		Use	Use	Use	Users
330.0	80	1	5	4	3
	60	3	16	15	11
	40	9	68	66	43
	20	18	202 *	231 *	246 *
330.1	80	1	1	2	3
	60	3	5	6	8
	40	7	14	19	20
	20	27	35	51	70
330.9	80	2	4	5	6
	60	5	10	13	13
	40	14	17	27	28
	20	22	34	53	55
338	80	0	1	1	1
	60	2	3	4	4
	40	4	11	12	11
	20	11	41	45	38
657/658	80	1	1	1	1
	60	2	3	4	5
	40	5	9	10	12
	20	11	29	27	39

(\* These 3 cases illustrate the 'Samuelson problem', ref. p. 183)

Table 4.7 Cumulative transaction threshold results for each user group by class Economics/Management for 1979.

It is apparent that 'new' thresholds in most cases are substantially greater than the 'old' thresholds and that they are of the same order over the range of transaction activity figures. Further, the fact that most of the thresholds for the 80 per cent level are 1 (i.e. the smallest possible value) shows that it would not be practical to "warehouse weed"<sub>15</sub> for transaction thresholds above 80 per cent in a one-year analysis.

Note the implications of analysing over only one year. It can be conjectured that there is little difference between two titles, one of which is used once in one year; the other being used once in the subsequent year. Exclusion or inclusion in an "80 per cent threshold one-year core" can thus be regarded as haphazard. The alternative side of the argument is that the costs of an error in classification are not great.

The resolution of the problem of what to do with low moving titles requires analysis over a longer time span than one year. This is the concern of Section 4.5.

An alternative to working with a percentage of cumulative transactions approach is to take a number of the most used titles at either the high or low end of the use distribution. This can be adequate and is efficient for certain problems such as determining the sufficiency of multiple copy holdings for titles that are heavily used.

It also has the practical merit of ensuring that the number of titles to be handled is manageable with the staff resources that are available.

Some relevant figures are:

Most Active Titles

1979 Economics/Management

	All	'Old'	'New'	'Total'
	Users	U/G	U/G	U/G
1000 Most Active Titles				
Actual # Titles	1031	983	939	957
Transactions Involved	23316	8611	11511	18544
% Total Transactions	56	58.9	69.7	59.6
% Active Titles				
of Total Titles	11.3	10.7	10.2	10.45
% Active Titles/Titles				
- 'O' use Titles	11.3	19.3	21.5	13.4

1980 Economics/Management

	All	'Old'	'New'	'Total'
	Users	U/G	U/G	U/G
1500 Most Active Titles				
Actual # Titles	1620	1752	1316	1503
Transactions Involved	21868	11287	14209	23207
% Total Transactions	44	67	69	62
% Active Titles				
of Total Titles	14	15	11	13
% Active Titles/Titles				
- 'O' use Titles	14	28	24	17

1979 Engineering

	All	'Old'	'New'	'Total'
	Users	U/G	U/G	U/G
1000 Most Active Titles				
Actual # Titles	431	549	438	516
Transactions Involved	7665	4635	3040	6825
% Total Transactions	43.0	62.0	65.7	56.3
% Active Titles				
of Total Titles	9.5	12.0	9.5	11.4
% Active Titles/Titles				
- 'O' use Titles	9.5	22.3	25.7	15.9

1980 Engineering

	All	'Old'	'New'	'Total'
	Users	U/G	U/G	U/G
500 Most Active Titles				
Actual # Titles	514	441	480	465
Transactions Involved	9490	4437	3675	7303
% Total Transactions	45.7	51.7	68.8	52.0
% Active Titles				
of Total Titles	10.1	8.7	9.4	9.7
% Active Titles/Titles				
- 'O' use Titles	10.1	15.3	26.0	12.5

Table 4.8 Percentage of transactions accounted for by the most active titles in the Economics/Management and Engineering disciplines for 1979/80.

The cases of Table 4.8 should be interpreted as follows.

In, for example, the table for engineering for 1980 there is an attempt to determine the 500 titles that are most heavily used by a particular user group, e.g., the 'old' undergraduate group. This is done by looking at the distribution of transactions against titles. As has already been mentioned (page 183) this distribution is not continuous. The closest we can get to 500 is in fact to consider the group of 441 titles where use in the particular year exceeds four transactions, this is Row 1 of the table.

Row 2 gives the number of transactions for those titles: 4437. This is 51.7 per cent of the total number of 'engineering' transactions by the 'old' undergraduate group in 1980. (This latter figure was 8576).

Now in that year all users borrowed 5077 titles. The 441 that we are looking at here is 8.7 per cent of that total. But not all of those titles were used by the 'old' undergraduate group; in fact 2188 were not used by them at all. Thus the 441 that we have represents 15.3 per cent of the (5077-2188) titles that were used by the group. This is the final row.

Using varying thresholds from Tables 4.5a and 4.5b the measures of equations 4.2, 4.3, 4.4 were evaluated for the 1979 and 1980 data, the results were as follows:

% of Transactions	Efficiency		Relevance		Overlap	
	'79	'80	'79	'80	'79	'80
80	74	75	100	100	29	28
60	67	66	97	97	26	24
40	64	66	90	94	26	25
20	51	53	90	94	21	20
Top 1500 Titles						

Economics/Management, Analysis of 1979 and 1980 Use.

Table 4.9a.

% of Transactions	Efficiency		Relevance		Overlap	
	'79	'80	'79	'80	'79	'80
80	78	100	100	90	25	20
60	78	83	97	94	17	16
40	71	76	98	95	15	14
20	74	59	93	80	6	15
Top 500 Titles	61	60	100	100	17	18

Engineering, Analysis of 1979 and 1980 Use.

Table 4.9b

Table 4.9 Efficiency, Relevance, Overlap analysis for the disciplines reviewed for 1979 and 1980.

A two-year analysis for 1979-80 was done for Dewey classes in Economics/Management and revealed no significant difference in the two sets of data. (Table 4.10)

% of Transactions		Efficiency		Relevance		Overlap	
		'79	'80	'79	'80	'79	'80
330.0	80	68	71	91	96	28	37
	60	40	41	88	85	20	22
	40	33	23	100	75	22	15
	20	50	100	100	100	50	100
330.1	80	75	76	100	100	30	26
	60	69	44	93	93	33	22
	40	55	45	92	83	30	27
	20	100	100	100	100	67	25
330.9	80	86	65	99	100	47	35
	60	82	62	99	100	46	37
	40	80	76	100	97	49	34
	20	69	64	100	90	38	36
338	80	70	76	100	100	27	29
	60	82	61	96	96	27	26
	40	58	64	94	99	22	26
	20	50	63	100	100	25	26
657-658	80	100	100	88	90	23	24
	60	70	69	96	96	21	20
	40	61	74	88	92	17	20
	20	48	63	79	92	15	12

Table 4.10 Efficiency, Relevance, Overlap analysis for the classes for 1979 and 1980.

#### 4.3.4 Allowing for Different Group Characteristics

The foregoing has been based largely on the notion that 'core' can be determined by setting a percentage of cumulative transactions and determining the group of most active titles that can account for that percentage.

It has been remarked that the threshold on title activity or the percentage of the cumulative distribution need not be the same for all user groups. Nonetheless, fairly comparable figures have been used in Section 4.3.3.

The following can be considered as proceeding along similar lines although now the variations between the user groups will be more marked. In effect the comparability between the groups will not now be based on cumulative transactions percentages. Alternatively, it may be construed that allowance is being made for the different characteristics of the different user groups.

Suppose one postulates that 'old' undergraduates are enrolled in subjects with small classes and that they specialise. Further, suppose that this specialisation may require the study of topics over long periods, (e.g. the period may be several weeks rather than the shorter interest spans encountered when the topics change weekly). Then it is possible that the demand for groups of items will extend over longer periods than with 'new'

undergraduates and will approach the characteristics of academic/postgraduate use. In such cases, the appropriate threshold for 'old' will be much lower than for 'new' undergraduate use. (The term 'appropriate' is used in the sense of servicing this type of use.)

It is worth recalling Kent's simple definition of core collection as:

"an item [can be considered] as being in the core collection if it has circulated externally one or more times thus being in the active circulating collection."<sup>17</sup>

Kent further suggested that "once an item enters the core collection it will always remain in that category."<sup>18</sup>

The notion of old undergraduate use can be handled by setting a threshold of "one" for this user group. In one sense, this may be considered as a Kent-type approach though it should be noted that it is being applied over one year whereas Kent uses the lifetime of the title.

The 'new' and 'total' thresholds continued to be determined on the basis of a percentage transaction.

Thus they maintain the values given in tables 4.5a and 4.5b. That is, the 'old' core included all items used by the 'old' undergraduate group in the year in question whilst the 'new' core was based on percentages of the transactions for that group. The core for total was also based on percentages of transactions.

The combination of 'new' and 'old' allowed for reflection of the different characteristics of the two types of use. The choice of a percentage for 'total' was in accord with the general view from the literature. This was expected to well reflect the needs of 'new' students whilst it would provide only a poor representation for the others. The only alternative was to take a threshold of, say "1" for the 'total' group. The inaccuracies now would have come from the 'new' group and would have been reflected in poor relevance figures.

The following results for the algorithms occurred where 'old' undergraduate thresholds were held at "1" while 'total' and 'new' were determined from a percentage specification.

% Transactions	Efficiency		Relevance		Overlap	
	'79	'80	'79	'80	'79	'80
80	74	75	100	100	29	28
60	45	45	100	100	20	20
40	14	17	100	100	7	9
20	3	4	100	100	1	2

Economics/Management 1979 and 1980.

Table 4.11a

% Transactions	Efficiency		Relevance		Overlap	
	'79	'80	'79	'80	'79	'80
80	78	100	100	90	25	20
60	46	45	100	100	16	12
40	19	19	100	100	8	6
20	5	5	100	100	2	2

Engineering 1979 and 1980.

Table 4.11b

Table 4.11 Efficiency, Relevance and Overlap figures for 'old' core with titles held at one transaction, and with 'new' and 'total' core being varied on percentage of transaction activity.

#### 4.4 Analysis, Comparison of Results for the Two Cases.

##### 4.4.1 Results for the Varying Threshold Approach

This relates to the results associated with Tables 4.5 and 4.9 and 4.10 which allowed thresholds to change. These results will be referred to as 'VTR' from here on.

##### Relevance

The relevance factor results for both disciplines indicated that relevance was high. In 1979/80 no relevance figure fell below 90 per cent. This implies that for every hundred titles in the 'total' core (i.e. the titles where all undergraduate use exceeds the 'total' threshold) nearly all will also be in the 'combined' core. In short, the allocation of titles for detailed analysis on the basis of total loan figures will provide a group of titles where 90 per cent of the titles would have been selected by a more detailed analysis. (There are advantages in the case of analysis and the comparatively 'low' requirements of the data needed for the 'total' analysis.)

It is also apparent from results for the classes (Table 4.10) that the "relevance" remains high when comparatively narrow parts of the collection are examined. There are only four cases where relevance is less than 90 per cent.

(These are associated with 80 per cent and 40 per cent transaction cut-off points in class 657-658, and in 40 per cent transaction cut-off points in 330.0 and 330.1.)

In two cases, the number of titles (386 for 330.1; 268 for 330.0) at the transaction percentage cut-off points are very low and establishing an effective core at such a low level has problems. These cut-off points are unlikely to provide a workable core since only very few titles will be extracted.

In engineering the relevance was likewise high.

### Efficiency

Taking the 80 per cent cut-off point in economics/management, the results (Table 4.9) show that 25 per cent of the titles in combined core will not be retrieved from the 'total' undergraduate core. This implies that 'total' undergraduate core includes many titles that will also be in 'combined' core but that there are some that are not selected on the 'total' basis.

The efficiency factor results for the two-year Dewey class analysis at the 80 per cent level produced similar results to those obtained for the Dewey disciplines (Table 4.10). There was only one case of efficiency dropping below 70 per cent (Dewey class 330.9, 1980).

It is important to note that the concepts "revelance" and "efficiency" as used in this research are "hard". That is, a title is either correctly or incorrectly classified. This permits the easy development of algorithms but in practice a misclassification is not disastrous. In determining some figure of merit for an approach, there is a case for weighting the errors. The way in which this should be done will depend on the application since it is the effect of a misclassification on the total exercise that is important. This point will not be followed up in this thesis. It should be noted that this observation is general and is not particular to any given way of determining the groupings.

#### 4.4.2 Results for the Group Characteristic Approach

This relates to the results given in Tables 4.11a and 4.11b where thresholds for 'old' undergraduates were held at "1". These results will be referred to as 'GCR' from here on in.

Note that Kent proposed that all active titles be in core. Here the requirements of two or more transactions in a year is more severe and it is felt it goes some way towards minimising the selection due to the random effects mentioned by Morse.<sup>19</sup> Note that the 'Kent view' may be expected to give rise to a greater combined core than the

threshold used here. Accordingly, the efficiency will be lower for the threshold of '0' (Kent) than here. One may also expect the figures for overlap to drop with the threshold used for the 'old' group.

Thresholds for 'new' and 'total' were selected as per Table 4.5; the "efficiency" and "relevance" factors were applied to the discipline user distributions.

The "relevance factor" as applied to the disciplines using 'GCR' remained high, achieving better results than with 'VTR'.

This is to be expected. The groups of titles for the 'new' and 'total' classifications have not altered while that for 'old' will have increased in size. The departure of relevance from 100 per cent is due to those titles which are in 'total' but which are not in the 'combined' core. Going from the 'VTR' to 'GCR' methods will effectively reduce the threshold for 'old' thus increasing the size of 'combined' core. It is likely that many titles in 'total' but not in the 'VTR' combined core will be in the 'GCR' combined core. Accordingly relevance rises.

The "relevance factor" obtained for all transaction cut-off points in economics/management was 100 per cent. This implies that titles retrieved by 'total' will all be

included in the 'combined core'. The figures for both disciplines are comparable. This result provides confidence in the use of 'total' as an indicator of titles that will be in 'combined core'. (It is, of course, now somewhat easier for a title to get into the 'combined core'.)

Efficiency, on the other hand, is significantly lower than in 'VTR'. At the 80 per cent transaction cut-off point efficiency is the same, however. In most other cases the result is less than half the previous figures.

The reasons for this are clear from the way in which the title sets change as outlined above. The efficiency measure has 'combined' core as the denominator and thus it may be expected to fall as the number of titles in 'old' rises.

For the 20 per cent transaction levels, efficiency drops to some 3 per cent. As indicated earlier a core that reflects a small number of titles will only be appropriate to certain library tasks, such as review of the adequacy of multiple copy holdings for the high movers. In such a case 'GCR' is clearly less appropriate than the 'VTR' procedure. Accordingly, the very low efficiency figures should not be construed as denoting that the 'GCR' approach is not useful. It is merely not appropriate in such situations.

Note that the efficiency factor in 'GCR' would be expected to be much lower since the pool of titles being considered in 'old' is much higher than that obtained with any threshold value of 'VTR'.

As indicated by the results, as many as 70 per cent of the titles in 'combined core' will not be retrieved by 'total'. This is completely different from the 'VTR' results.

The results indicate that the use of 'total' to look at the behaviour of various undergraduate groups is adequate except when the characteristics for user groups are required to be similar. If this is not the case then it can reasonably be questioned how the total transactions are to be categorised. Will the requirement follow the new group or the old? When reflecting on this, note the important point that with 'total' approach only the total loan figures are available. It is not possible to postulate a categorisation rule that reacts differently to loans for each group. The group figures will not be available.

The decision as to which version and methodology to use in examining library use depends on the purpose of the analysis. 'VTR' is designed to look at high and low usage titles with the intention of providing an indication of which titles need to be held and the number of copies that

may be required. It provides an analysis based on transaction levels and takes the differences between 'new', 'old' and 'total' undergraduate use.

In circumstances where the 'VTR' approach is appropriate and where there are comparable percentages for each group, then the results of this thesis suggest that it would not be necessary to establish 'new' and 'old' cores separately. However, where it is desirable that there be significant differences between the characteristics sought for various user groups, then it is apparent that a 'total' approach may have significant shortcomings.

#### 4.4.3 The Location Problem

We now look at the location of titles within a library operation composed of a number of sub-libraries with significantly different roles. This discussion can embrace the relegation problem where there is closed access to part of the collection, the high use/low use arrangement and the undergraduate/research classification. The way in which the various approaches are used depends on the overall aim for the various sub-libraries.

'VTR' does not give an indication of where to hold this material. If the library policy is to separate the low and high use titles of the collection the 'GCR' approach has advantage if the data is available. 'GCR' establishes

a core transaction level that effectively includes all material borrowed for any reason and is more appropriate for solving this type of question.

The relevance and efficiency factors provide an indication of overlap between the 'total core' and 'combined core' but not between 'old', 'new' and 'total' cores. If the collection is to be divided into high and low use material it is necessary to look at the overlap in the use of titles by the sub-groups under review.

If looking at overlap of titles between 'old' and 'new' core there is no difficulty in determining the titles that will be common to both. This basic overlap is discussed in relation to equation 4.1:

$$\begin{array}{rcl} \text{OVERLAP} & = & \frac{\text{TTCM1}}{\text{TOLD} + \text{TNEW} - \text{TTCM1}} \\ (\text{OLD/NEW}) & & \end{array}$$

(4.1)

It is more convenient to be able to use total use figures than to have to consider the user subgroups. Relevance and efficiency factors are based on 'total' undergraduate use and 'combined core' and these provide two measures of the adequacy of this simplification. Here we look at overlap between the various 'cores' and, in particular, express the results in terms of total.

Ratio 1

This gives the overlap of titles between the three categories 'old', 'new' and 'total' undergraduates as a percentage of 'total' undergraduates.

This is:

$$\begin{array}{lcl} \text{OVERLAP} & = & \frac{\text{TTCM2}}{\text{TOTAL}} \\ (\text{OLD, NEW, TOTAL}) & & \end{array}$$

(4.4)

In Venn diagram form this can be represented as:

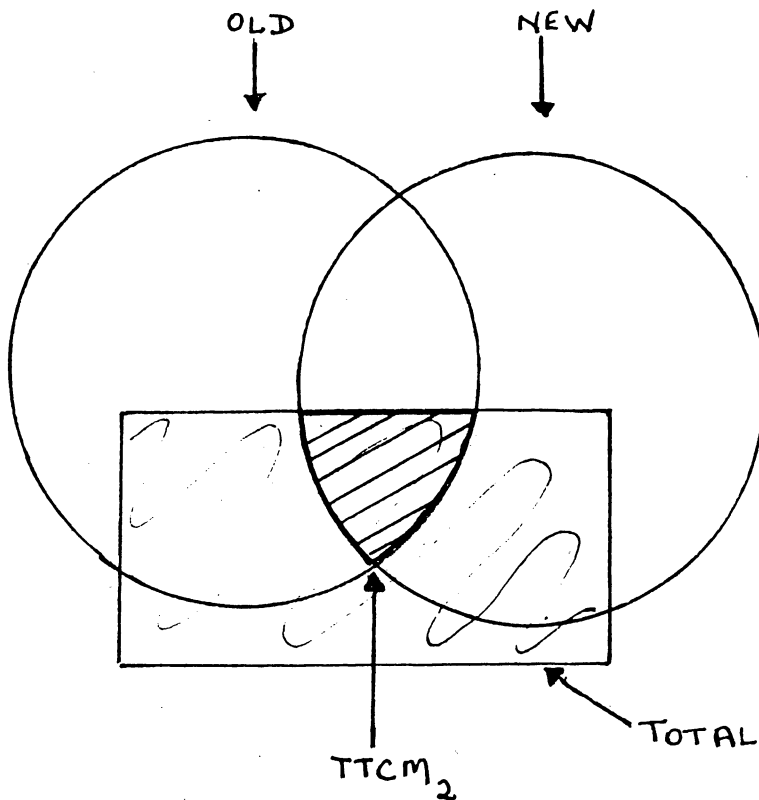


Figure 4.4. Overlap as defined in Equation 4.4

Table 4.12 gives results for Ratio 1 and compares it with the value for overlap which is of more direct concern, viz, overlap (old/new).

If the overlap figures are high then the decision on location of an item may not be straightforward. There could even be a case for holding at least one copy in each sub-library where there is activity. Alternatively, if there is a low overlap as in Table 4.12, then the location task is simplified.

The overlap calculation is particularly pertinent to the relocation task (as opposed to the initial allocation of a title). Note that the groupings to be used need to be determined on the basis of the aims of various sub-libraries. Here the comparison was between user groups (viz, 'old' and 'new' undergraduates). This is well suited to a case where the aims of a sub-library are related to a readily identified user group.

Note also that as the loan periods for various sub-libraries may differ, the overlap calculation can also be relevant to determining loan policies.

Table 4.12 shows that in Engineering the overlap tends to reduce with the percentage of total transactions. This is reasonable and it reflects a difference between the 'high-movers' for the two user groups. When the

transaction threshold is low, attention is focused on these high-movers; when it is high, a larger title population is captured. It is reasonable to expect that a high-mover for one user group will be a moderate-mover for the other. Thus the overlap figures are not too low where the percentage of transactions captured is high.

Another pattern that could occur would be one where the overlap is mainly a reflection of subject interest. In such a case there would tend to be more uniformity in the overlap measure. (This may have been the case with the 'VTR' calculations for economics/management).

Ratio 1 has a denominator that is the number of titles in 'total'. This will be a function of the threshold set for the cumulative transactions. The numerator is the number of titles common to all three groups, 'new', 'old' and 'total'. In the case of 'VTR' it is reasonable to expect that the dependence on transaction threshold will be comparable between Ratio 1 and old/new overlap. For 'GCR', however, there will be significant differences between the denominators for the two measures. In the GCR approach there will be an important dependence on the 'old' group of the combined core. It may be expected that the combined core is near constant in composition whilst the set of titles common to both user groups will reduce with the percentage of transaction capture sought. Thus as that percentage reduces, so too does the old/new

overlap due in the main to the reduced number of 'new' titles. If, however, Ratio 1 is used then there is scope for the numerator and denominator to reduce together. The merits of Ratio 1 for the GCR cases are not too obvious.

Table 4.13 gives some overlap results for the VTR cases for the classes. The number of titles in the 'total' set is also given in each case. Only the values for the VTR approach are given.

It may be noted that the overlap tends to be class dependent; this is not surprising. There tends to be a constant relationship between the overlap measure and Ratio 1 within a class.

The figures for the population size indicate that the values determined need to be treated with some caution. Mention has been made elsewhere about the difficulty in determining a value of transaction/title that will give rise to a figure of precisely 80 per cent of the cumulative transactions.

Even in the VTR approach there may be inconsistency between the 80 per cent for the 'old', 'new' and 'total' groups. In short, there may not be a comparison between like values that is sought. In the case of the class analysis, this problem becomes greater because of the very small groups of active titles. Consider, for example, the

implications of a sensitivity study of the class 330.0 for the approach VTR20.

The values in Table 4.13 are useful for the general understanding that they give of the situation. They should not be misapplied by reading too much into them. It pays to recall their precision.

Ratio 1

Approach	Economics/Management		Engineering	
	Overlap	Ratio 1	Overlap	Ratio 1
VTR 80	29	39	25	32
VTR 60	26	38	17	22
VTR 40	26	36	15	20
VTR 20	21	37	6	7
GCR 80	29	39	25	32
GCR 60	20	44	16	34
GCR 40	7	51	8	38
GCR 20	1	51	2	36

Table 4.12 Overlap and Ratio 1 for disciplines and approaches 1979.

<u>Class</u>	<u>Approach</u>	<u>Overlap</u> (Old/New)	<u>Ratio 1</u>	<u>Total</u>
330.0	VTR 80	28	38	76
	60	20	44	26
	40	22	67	3
	20	50	100	1
330.1	VTR 80	30	40	131
	60	33	44	41
	40	30	50	12
	20	67	67	3
330.9	VTR 80	47	54	161
	60	46	55	71
	40	49	61	28
	20	38	56	9
338	VTR 80	27	38	840
	60	27	31	305
	40	22	36	85
	20	25	50	16
657-658	VTR 80	23	20	960
	60	21	25	386
	40	17	17	121
	20	15	25	28

Table 4.13 Overlap and Ratio 1 for a range of classes and approaches in the Economics/Management area.

Ratio 2

The previous discussion has been largely concerned with the use of 'total' (as variously defined) to estimate 'combined core'.

Overlap problems do not necessarily arise from 'total', however, and they may be intrinsic to the choices made by the users. Equation 4.5 looks at this problem and measures overlap of titles as a percentage of combined core.

$$\text{OVERLAP} = \frac{\text{TTCM}_2}{\text{TOLD} + \text{TNEW} - \text{TTCM}_2} \quad (4.5)$$

In Venn diagram form this can be represented as:

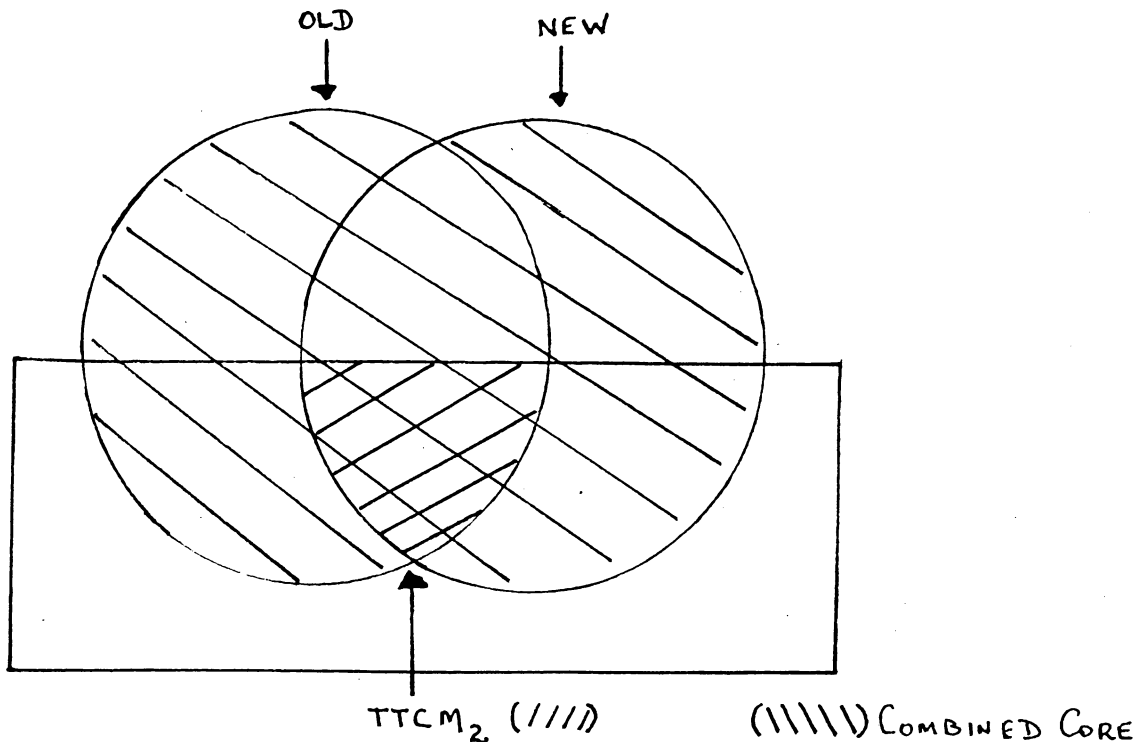


Figure 4.5 Overlap as defined in Equation 4.5

A little manipulation shows that Ratio 2 may be determined from other values. Note that Ratio 1 and Ratio 2 have the same numerator. This also applies to relevance and efficiency.

$$\frac{\text{Ratio 1}}{\text{Ratio 2}} = \frac{\text{Combined Core}}{\text{Total}} = \frac{\text{Relevance}}{\text{Efficiency}}$$

(4.5)

#### 4.5 Long-term Analysis

This section presents the results from analysing the behaviour over several years. The years reviewed were 1972, 1973, 1976, 1978, 1979, 1980. The intention was to examine all years from 1972-1980. However, corruption of the data did not permit this.

The long-term analysis of use extracted cases which had maximum loans over the period in any one year and looked at the behaviour over the following years. For example:

	Year					
Call Number	1972	1973	1976	1978	1979	1980
330/1	9	38	40	16	14	19

Here the maximum usage was 40 transactions and this occurred in 1976. The long-term analysis would look at behaviour in 1978, 1979, and 1980 and compare this with activity in 1976. The aim was to see whether there are use patterns that may assist in exercises such as culling.

The analysis looks at the falling-off of activity after the maxima. It is also important to determine whether the pattern is cyclical and whether activity consequently is likely to rise once a local maximum is passed. (The maxima may not be exceeded by the initial fall-off but may be reversed by a subsequent rise).

Determination of these patterns would provide some indication as to how long it is necessary to wait before a culling exercise can be undertaken without removing material that may be required in later years.

Note the relevance of the size of the maxima to the culling exercise. There is considerable effort (staff-hours) required to make a decision on a title. If the possible return from such an exercise is a reduction of the number of copies by one, then the exercise is very costly. However, the removal of five of the copies of a title may be worthwhile from a cost-benefit point of view.

To appreciate the point about the size of the maximum use consider two situations. In the first, if the maximum use is three in one year and it falls to one or two in the following years, the drop does not reveal any pattern and loans are really at a 'noise' level. Alternatively, if the loan level falls from 30 loans per annum to 10, the difference can be important when culling.

The initial processing was designed to provide a broad analysis of the number of cases in any one year which had a maximum value in that year.

The results were:

Year	1972	1973	1976	1978	1979	1980
Cases	1272	1708	2277	2440	2645	5261

These results relate to maxima for undergraduate loans only. Note that the 1972 and 1980 cases may not represent true maxima; it is not possible to compare with activity in adjacent years. Likewise 1973, 1976 and 1978 have their problems due to the one-sided nature of the data.

#### 4.5.1 Distribution

To provide a detailed analysis, a cumulative distribution was generated for activity of titles and transactions. The activity was examined on the basis of maximum use titles for all undergraduate use over the six years reviewed.

The distribution was:

Number of Transactions	Titles	Cumulative Titles	Cumulative Transactions
572	1	1	572
514	1	2	1086
399	1	3	1485
294	1	4	1779
⋮	⋮	⋮	⋮
126	1	35	7302
125	2	37	7552
⋮	⋮	⋮	⋮
36	6	263	21172
35	9	272	21487
⋮	⋮	⋮	⋮
4	1191	5168	63303
3	1830	6998	68793
2	2944	9942	74681
1	5661	15603	80341

Table 4.14 Sample of cumulative distribution for  
title/transaction activity for maximum use  
title over all years for all undergraduate use.

The analysis revealed typical diminishing return behaviour, e.g. a high number of titles with a maxima activity of one use and a few titles with high maxima.

It should be emphasised that the cumulative transactions above are not the transactions in any one year. They are formed from the distribution of Table 4.9

$$\sum_{\text{titles}} \quad (\text{maximum transactions})$$

To provide a detailed view across a broad range of activity, the cumulative distribution table, set at various maximum transaction levels ranging from 80 to 20 per cent, was established.

The results were:

Cumulative % Transactions	Maximum Transaction Level	Number Titles
80	4	5168
60	8	2145
40	18	703
20	55	145

Table 4.15 Cumulative transaction threshold results for maximum transactions for the period under review for full undergraduate use.

This table provides the basis for deciding on the level of maximum transaction use which will provide some return on the effort expended during a culling exercise.

At the 80 per cent level the number of titles that would need to be reviewed is 5168, and the maximum transaction level is 4. This level of maximum use is very low and any analysis that looked at drop-off rates in the following years would have difficulty in distinguishing signals in the 'noise'. It would be difficult to decide whether the reduction was a true drop-off or whether the whole period

was 'noise' use. Apart from this, however, a great deal of effort would be expended with very little return as the number of copies expected for titles with that level of maxima activity would be low. At the other end of the distribution, 20 per cent would provide 145 titles for review, too few to provide a significant return.

Any culling exercise should offer the probability of removing a significant number of copies if the exercise is to be worthwhile. It should avoid discarding material that may be of use in the following years.

The transaction level chosen for further analysis was 40 per cent; this accounted for 31,899 of the cumulative transactions and provided 703 titles with a maximum transaction level of 18.

In addition to establishing a relevant cut-off point for transaction levels, it was also necessary to determine a relevant fall-off point. The literature does not appear to discuss what is a significant percentage reduction in use over a number of years. For example, a drop from 18 to 15 in the subsequent year represents a 16 per cent fall, but the number of copies required to support both levels of activity is comparable. Further, such a small drop does not provide confidence that use has, in fact, fallen off sufficiently that material is no longer required in the same numbers as previously.

For the present purposes, a level of fall-off that was considered significant was 60 per cent.

This was large enough to avoid the problems that may occur from 'noise'. Recovery from this level of drop-off would also indicate any cyclical patterns if there were rises in any subsequent years.

#### 4.5.2 Maximum Transaction Level Analysis

In choosing the maximum transaction level of 18, the initial processing established the number of cases in each year with this maximum.

The results were:

Year	1972	1973	1976	1978	1979	1980
Cases	123	133	94	122	83	148

Table 4.16 Title/year distribution of cases for all undergraduate loans with a maximum transaction level  $\geq 18$  for the total period under review.

It should be emphasised that the 1980 and 1972 cases could not be examined since it was not possible to ascertain whether the cases had, in fact, reached a maximum since it was the last year in the period reviewed.

The analysis then focused on 1978 maxima with use  $\geq 18$  since this permitted the determination of the distribution of use ratios for a further two years. (This is also a little unclear since it is not possible to determine whether the maxima occurred in 1977 or 1978.)

Two ratios were calculated:

- . The first  $R_1$  calculated 1979 use as a percentage of 1978 use and established the percentage by which use fell.
- . The second  $R_2$  calculated 1980 use as a percentage of 1978 use and established the percentage by which use fell.

For example:

Call number	1978	1979	1980 (Use)
300/269	24	14	15

$$R_1 = \frac{14}{24} = 58\%$$

This indicates that use fell to 58 per cent or, alternatively, the percentage fall in use was 42 per cent.

$$R_2 = \frac{15}{24} \quad \text{or} \quad 62.5\%$$

This indicates that from 1978 to 1980 use fell to 62.5 per cent or, alternatively, the percentage fall in use was 37.5 per cent.

It was considered more appropriate to look at cases where use decreases from 1978 to 1979 but there is no guarantee that it will continue to be low in 1980. Although the local maximum appears to be in 1978 since 1979 activity is lower than in 1978, the maximum need not necessarily be in 1978 since it is not known whether 1980 will become the maxima. (There may be a problem with cyclic data, as was mentioned earlier.)

There were 138 cases where the maxima in 1978 was greater than or equal to 18.

Figure 4.6 gives a scatter diagram for  $R_1$  and  $R_2$  for titles having maximum undergraduate use in 1978 for the period 1972-1979.

A review of the scatter diagram indicates that if a 60 per cent drop in use is required from 1978 to 1979, that is an  $R_1$  value equal to or less than 40 per cent, the number of titles that would qualify is 37. If weeding was

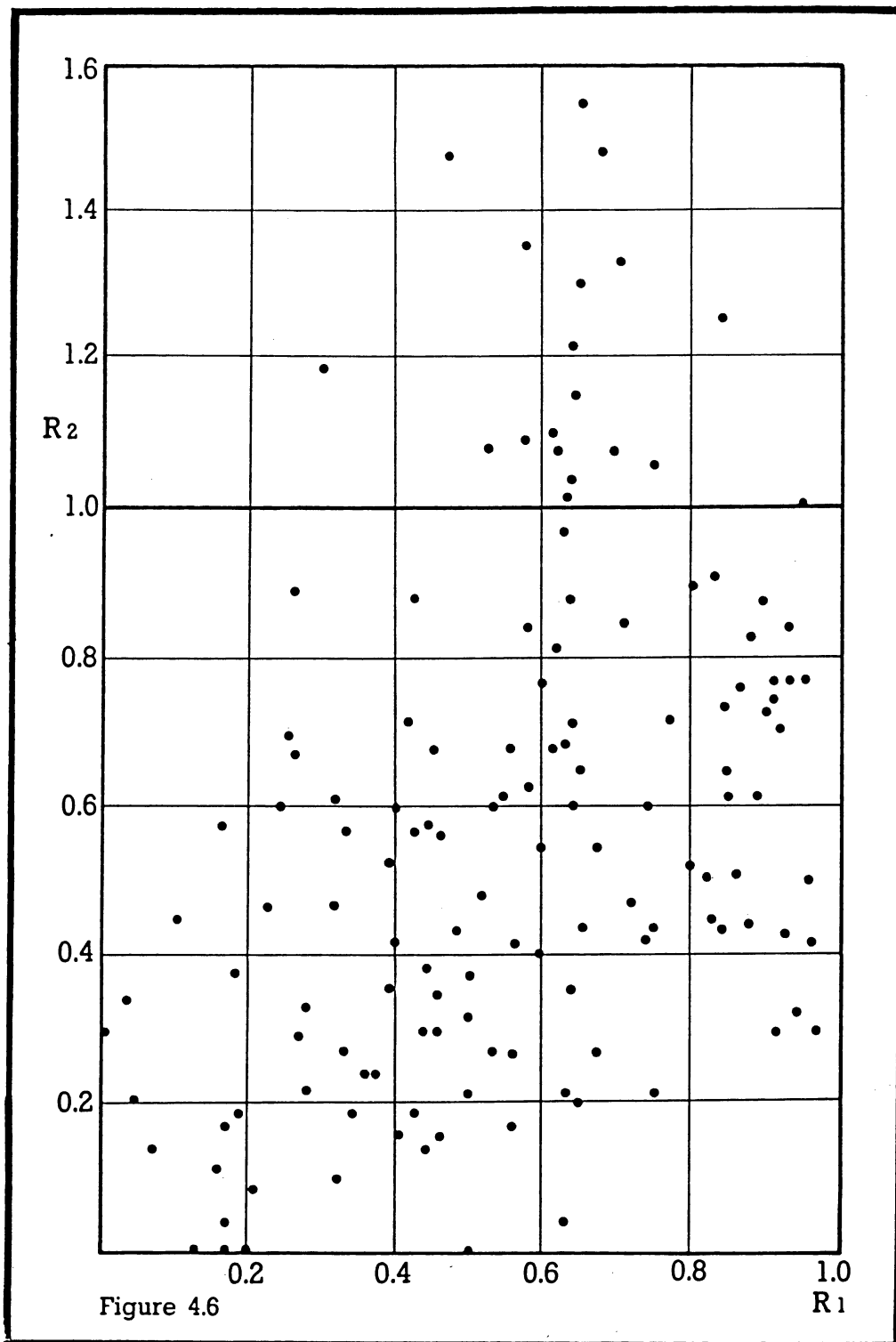


Figure 4.6. Scatter diagram for  $R_1$  and  $R_2$  for titles having maximum undergraduate<sup>2</sup> use in 1978 for the period 1972-1979.

$$R_1 = \frac{U/G \text{ '79}}{U/G \text{ '78}} \quad R_2 = \frac{U/G \text{ '80}}{U/G \text{ '78}}$$

undertaken on the basis of the 1979 percentage fall-off rate, 37 titles would have been examined.

Further examination of the scatter diagram is worthwhile.

If it is required that the same 60 per cent drop-off be maintained in the second year, 1980, (that is an  $R_2$  value equal to or less than 40 per cent), only 23 titles would meet the criterion in the second year.

This shows that only 23 titles will have maintained the required percentage fall to be weeded with confidence.

If all 37 titles had been weeded and the number of copies had been reduced, 14 titles would have had a copy deficiency (nine titles in the second year had a fall of between 59 and 40 per cent, and five between 30 and 0 per cent). The single title with an  $R_2$  of 1.2 indicates that in 1980 the title rose rather than fell. Although it is not possible to ascertain whether this 1980 use is a maximum, the weeding of this title which is obviously on an upswing would have been in error.

There are a number of points that need to be made.

In looking at high movers, that is maxima of  $\geq 18$ , the pool retrieved for culling purposes was quite small, 138 titles. In any culling exercise the effort expended is only warranted if the probability of reduction in the

that weeding of 138 titles would make available would not be large. It is apparent from the results that the threshold would need to be dropped to provide an adequate pool of titles; this next step was not carried out in this thesis.

The view of some publications that fall-off occurs quickly was not supported by the data presented here. In most cases there is a gradual reduction.

Although the data indicates that once use has fallen-off it is likely to remain down, this cannot be proven conclusively without several more years' data which would detect cyclical trends with more precision.

It is not suggested that titles at an  $R_2$  of 18 and 19 cannot be weeded. What is suggested, however, is that in order to weed material at these ratios, it would be necessary to wait for more than two years since weeding within that time could produce serious error.

Effectively, it will take several years for use of those titles to fall to a figure of say 60 per cent maximum.

#### 4.6 In-House Survey Results

This section presents the results obtained from the three week in-house survey undertaken in the University of New South Wales library. It outlines a rudimentary cluster analysis that is by no means exhaustive. The results are intended to provide a basis for future work in the development of a 'theory of clusters' for in-house use analysis in academic libraries.

##### 4.6.1 Survey Limitations

This in-house use analysis gives some reasons for making proposals about future work.

The data collection in the study was designed specifically to assist in the analysis of in-house use. Accordingly, one cannot ask more of the data with any guarantee of success. Studies with different roles would need to consider other methods of data collection.

The method of sampling needs close control with particular attention being directed to the effects of time-lags in the reshelving of material used in-house. If there is insufficient control here there may be major problems with duplicated samples.

The interval in which sampling occurs is also critical. A period of three weeks was chosen here and was considered representative, as discussed earlier in Chapter 2.

However, in other circumstances it may be essential to consider at least one full academic year. Such a time frame would permit studies of features that have not been discussed in this analysis, e.g. borrowing prior to examinations, borrowing prior to assignments or vacation.

In a centralised system, i.e. where material is returned to one central sorting area and then transferred to appropriate distribution points for shelving, the problem of control over repeated samples becomes a major problem. In this type of system it is not easy to indicate movement patterns across collections unless the pickup system is monitored.

#### 4.6.2 Cluster Analysis

To determine various borrowing patterns it was essential to develop a system that would look at alternative combinations of use.

A modified cluster analysis methodology was considered to be the most appropriate method for discerning the required patterns and for determining the eligibility of titles within specified categories of use.

The major categories of use here were intermittent versus concentrated use.

The latter required elaboration and the type of activity that needs to be considered in this category included:

- \* many copies being used over a short period of time
- \* very active single copies of a title
- \* frequent use of a title over a period of time
- \* several uses of a title in one day

To examine this activity, it was essential to set up cluster patterns that would approximate the activity patterns above.

The decision as to whether a group of events constitutes a cluster is a difficult one. It is a standard computing exercise but, for the purpose of this thesis, it is not necessary to discuss the mathematics of cluster analysis.

Any future in-house studies would find it essential to first define what was meant by 'cluster'. Such precision, was not considered essential in the illustrations of this thesis. To develop such definitions and to apply them would need data that is more accurate than the present and, as stated earlier, the cluster analysis used here indicates alternative use patterns only.

DAY

Title/ Copies	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294
330/1A							*				*		*	*				*			*	*			
330/1B	*									*	*				*		*			*		*			
330/1C	*								*		*				*			*		*					
330/1D	*							*							*			*		*			*	*	
330/1E		*					*				*			*		*	*		*			*			**

Table 4.17 Sample in-house use patterns representing cluster types 1-6

The use of the title as illustrated on the previous page is proposed to show concentrated use as per the patterns considered in the following discussion.

(i) Type 1 -- Several copies used in one day

Examples of this type of use occurred on days  
271 (Copies B, C, D), 280 (A, B, C), 284 (B, C, D),  
287 (A, C, D), 289 (B, C, D), 291 (A, B, E).

Any title which has more than one copy in use on one day is considered to have an occurrence of this type of cluster. Each day exhibiting this type of cluster is considered as one occurrence.

One title can, therefore, have more than one occurrence over the sample time, e.g. in the example above the title would be counted only once but the number of occurrences in this category total six.

The hypothesis in this case was that use of this type was reference use by a group. Further it would have to be established whether titles displaying this type of use have a copy(ies) located in limited access reserve collections or whether the demand in-house was a result of a copy(ies) not being in such a collection.

(ii) Type 2 -- One copy only in use on a given day

From the example, use on day 294 (Copy E) would qualify as an occurrence if the use recorded was for one copy only. This could indicate use of the one copy at different times on the same day or on different floors on the same day.

The collection of data in this category must be done carefully if there is to be any statistical relevance. In an in-house survey it is extremely difficult to determine more than one use in one day of a particular copy unless the sample is carefully monitored. Attention needs to be given to the effect of errors in data collection.

(iii) Type 3 -- One copy use over several days (with no date gaps)

The use on days 285 and 286 (Copy E) and 292 and 293 (Copy D) would be considered here, provided the use on both days was the same copy.

The detection of occurrences of this type could be important as it was conjectured that it may indicate use by one student only or it may indicate the unavailability of other copies and hence concentrated use of the one copy. Occurrences of this type,

lasting more than five days, could require analysis of circulation use of the title or examination of the reserve use.

- (iv) Type 4 -- A title with three or more uses over seven days

The loan period for undergraduates was one week. This category was established on the same duration as the loan period.

Three or more uses was chosen as relevant. Two or fewer uses may not indicate concentrated use but could be better thought of as intermittent use.

Reshelving and pickup delays needed to be incorporated and for the purpose of this thesis three uses, allowing for these delays, was considered to represent concentrated use.

From the example, use between days 276 and 282, days 283-290 would be considered occurrences in this category.

In this case the title would be counted once; the number of occurrences would be two.

- (v) Type 5 -- Several days use of a title with no more than a 3-day interval between uses

This category was established to provide an analysis that would not place any time limit on intensive use. To be considered intensive use, the time span between each use was to be between one and three days. More than three days was considered not to warrant inclusion as intensive.

From the example, the following would qualify as occurrences in this type of cluster:

Day 276 through to 294, as there is no period between where use is more than three days apart.

- (vi) Type 6 -- One copy use only with at least a 3-day interval between uses

One title could exhibit this type of use more than once. For example, in Table 4.17 use between days 272 and 280 for Copy E would be considered to be three separate occurrences of this type.

It is hypothesized that this was a single reference use at each occurrence or, alternatively, the item was not perceived as being sufficiently valuable for borrowing.

#### 4.6.3 General Conditions re Cluster

Many titles would qualify as occurrences in more than one type of cluster as shown from the example above.

In cases where this has occurred, the title has had occurrences counted in each category for which it is eligible.

Titles may qualify more than once for a particular category as shown in the Type 4 example. In such cases, the title is counted once and the occurrences are counted as many times as are eligible.

Finally, once a cluster pattern is established, the use continues in the cluster until the cluster pattern is broken.

#### 4.6.4 Results of In-House survey

The data was not sufficient to test a number of things conclusively. However, a number of conjectures can be made that would require a more controlled data collection.

(i) Conjectures

- a. Use of material on 'either side' of a particular title, indicates either selection or rejection of a number of alternatives by a number of students or by one student only. Caution is necessary in determining what constitutes 'either side'. If shelf-listing occurs on receipt of a title and the incorporation of the title within the collection varies between sub-libraries, 'either side' need not be identical in each sub-library.
- b. Use of the same material in more than one sub-library could indicate saturation in one collection and movement into another. It is not possible with the data collected here to assess this.
- c. The pattern of use in-house does not vary significantly in either the hard or soft sciences.

In this sample, the chosen areas showed predominantly intermittent use of one copy. (Type 1).

Breakdowns of titles/occurrences results appear in Tables 4.18 a - d, for each type of cluster use by Dewey class.

These tables have been established on a class basis and divided by the collections 'undergraduate', 'physical sciences' and 'social sciences'.

The sample on a Dewey section basis provided the following results:

Single Copy -- Intermittent Use

<u>Dewey Section</u>	<u>% of Use</u>
330 - 339.99	93.5
657 - 658 inclusive	88
620 - 629 inclusive	93

On the basis of these results, it would have to be assumed that most in-house use is single copy intermittent use of titles.

A lot of single use was expected due to the shelving delays discussed in Chapter 2. However, the results do indicate that multiple use was occurring in-house and this is the area that requires follow-up in a future study.

In the following cluster tables the abbreviations are:

1.    U       Undergraduate Library
- P       Physical Sciences Library
- S       Social Sciences Library
2.    1       Type 1 -- Several copies used in one day
- 2       Type 2 -- One copy only in use on a given day
- 3       Type 3 -- One copy use over several days (with  
                                no date gaps)
- 4       Type 4 -- A title with three or more uses over  
                                seven days
- 5       Type 5 -- Several days use of a title with no  
                                more than a 3-day interval between  
                                uses
- 6       Type 6 - One copy use only with at least a  
                                3-day interval between uses
3.    T       Number of titles
- E       Number of occurrences as explained on Page 243

330-334.99

TYPE OF CLUSTER

Class	1	2	3	4	5	6
	U P S	U P S	U P S	U P S	U P S	U P S
330-330.999	T 131 - 22 E 146 - 22	T 5 - - E 11 - -	T 4 - - E 4 - -	T 1 - - E 1 - -	T 6 - - E 8 - -	T 9 - 2 E 9 - 2
	U P S	U P S	U P S	U P S	U P S	U P S
331-331.999	T 80 - 66 E 85 - 66	T - - - E - - -	T - - - E - - -	T 2 - - E 2 - -	T 2 - - E 2 - -	T 2 - 1 E 2 - 1
	U P S	U P S	U P S	U P S	U P S	U P S
332-332.999	T 46 1 33 E 47 1 36	T 2 - 1 E 2 - 1	T - - - E - - -	T 1 - - E 1 - -	T - - - E - - -	T - - 1 E - - 1
	U P S	U P S	U P S	U P S	U P S	U P S
333-333.999	T 19 21 21 E 20 22 24	T 1 1 - E 1 1 -	T - - - E - - -	T - - - E - - -	T - 1 - E - 1 -	T - 2 - E - 2 -
	U P S	U P S	U P S	U P S	U P S	U P S
334-334.999	T 1 - - E 1 - -	T - - - E - - -	T - - - E - - -	T - - - E - - -	T - - - E - - -	T - - - E - - -

Table 4.18 Class analysis of the title/occurrences for each type of cluster described in Section 4.6.

335-339.99

TYPE OF CLUSTER

	U P S	U P S	U P S	U P S	U P S	U P S
335-335.999	T 46 - 11 E 56 - 12	T - - - E - - -	T - - - E - - -	T - - - E - - -	T - - - E - - -	T - - 1 E - - 1
	U P S	U P S	U P S	U P S	U P S	U P S
336-336.999	T 11 - 16 E 11 - 17	T - - - E - - -	T - - - E - - -	T - - - E - - -	T 1 - - E 2 - -	T 1 - - E 1 - -
	U P S	U P S	U P S	U P S	U P S	U P S
337-337.999	T 3 - 3 E 3 - 3	T - - - E - - -	T - - - E - - -	T - - - E - - -	T - - - E - - -	T 1 - - E 1 - -
	U P S	U P S	U P S	U P S	U P S	U P S
338-338.999	T 193 4 145 E 205 4 158	T - - 1 E - - 1	T 5 - - E 5 - -	T 5 1 2 E 5 1 2	T 5 - - E 5 - -	T 10 - 2 E 11 - 2
	U P S	U P S	U P S	U P S	U P S	U P S
339-339.999	T 23 - 38 E 30 - 39	T 4 - - E 6 - -	T - - - E - - -	T 6 - - E 6 - -	T 4 - - E 5 - -	T 5 - - E 6 - -

Table 4.18a

620-624.99

TYPE OF CLUSTER

CLASS	1	2	3	4	5	6
	U P S	U P S	U P S	U P S	U P S	U P S
620-620.999	T 52 35 1 E 58 35 1	T 2 - - E 2 - -	T 1 - - E 1 - -	T - - - E - - -	T 2 - - E - - -	T 9 - - E 9 - -
	U P S	U P S	U P S	U P S	U P S	U P S
621-621.999	T 103 118 - E 121 112 -	T 3 1 - E 3 1 -	T - - - E - - -	T - 1 - E - 1 -	T 6 - - E - - -	T 2 6 - E 2 6 -
	U P S	U P S	U P S	U P S	U P S	U P S
622-622.999	T 10 13 - E 12 18 -	T - - - E - - -	T - - - E - - -	T - 1 - E - 1 -	T - - - E - - -	T - 1 - E - 1 -
	U P S	U P S	U P S	U P S	U P S	U P S
623-623.999	T 1 - - E 1 - -	T - - - E - - -	T - 6 - E - - -	T - - - E - - -	T - - - E - - -	T - - - E - - -
	U P S	U P S	U P S	U P S	U P S	U P S
624-624.999	T 67 75 - E 75 83 -	T 3 1 - E 3 1 -	T 1 - - E 1 - -	T 2 2 - E 2 2 -	T 4 - - E 4 - -	T 6 5 - E 6 5 -

Table 4.18b

625-629.99

TYPE OF CLUSTER

	U P S	U P S	U P S	U P S	U P S	U P S
625-625.999	T 10 17 - E 11 17 -	T - - - E - - -	T - - - E - - -	T - - - E - - -	T - - - E - - -	T - - - E - - -
626-626.999	U P S T - - - E - - -	U P S T - - - E - - -	U P S T - - - E - - -	U P S T - - - E - - -	U P S T - - - E - - -	U P S T - - - E - - -
627-627.999	U P S T 9 7 7 E 10 7 7	U P S T - - - E - - -	U P S T - - - E - - -	U P S T - - - E - - -	U P S T - - - E - - -	U P S T - 1 - E - 1 -
628-628.999	U P S T 12 32 - E 13 37 -	U P S T - - - E - - -	U P S T - - - E - - -	U P S T - 4 - E - 4 -	U P S T - - - E - - -	U P S T 1 1 - E 1 1 -
629-629.999	U P S T 15 21 - E 15 22 -	U P S T - - - E - - -	U P S T - - - E - - -	U P S T - - - E - - -	U P S T - - - E - - -	U P S T - - - T - - -

Table 4.18c

657-659.999

CLASS	1	2	3	4	5	6
	U P S	U P S	U P S	U P S	U P S	U P S
657-657.999	T 46 - 34 E 70 - 38	T 10 - 4 E 15 - 4	T - - - E - - -	T - - - E - - -	T 7 - 4 E 11 - 4	T 11 - 6 E 15 - 6
	U P S	U P S	U P S	U P S	U P S	U P S
658-658.999	T 110 - 162 E 117 - 171	T 2 - 3 E 2 - 3	T - - - E - - -	T - - - E 1 0 0	T 2 - - E 2 - -	T 5 - 2 E 5 - 2
	U P S	U P S	U P S	U P S	U P S	U p S
659-659.999	T - - 1 E - - 1	T - - - E - - -	T - - - E - - -	T - - - E - - -	T - - - E - - -	T - - - E - - -

Table 4.18d

<u>Dewey Class</u>	<u>Titles</u>	<u>Trans</u>	<u>1 Use Only</u>
330 -330.09	40	77	34
330.1-330.99	123	190	88
331 -331.99	153	175	134
332 -332.99	84	100	70
333 -333.99	62	78	55
334 -334.99	1	1	1
335 -335.99	51	68	38
336 -336.99	27	34	22
337 -337.99	7	8	7
338 -338.99	341	472	261
339 -339.99	54	111	34
657 -657.99	97	220	52
658 -658.99	299	414	229
	1038	1948	1025
620 -629.99	90	119	71
621 -621.99	222	274	193
622 -622.99	23	34	16
623 -623.99	1	1	1
624 -624.99	138	216	94
625 -625.99	28	29	28
626 -626.99	-	-	-
627 -627.99	17	20	15
628 -628.99	44	78	33
629 -629.99	37	40	34
	600	811	485

Table 4.19 In-house survey results by class for titles, transactions and one use only titles.

Dewey Sections

	330-339.99	659-658.99	620-629.9
Titles with one use only as % of total titles	78.8%	70.9%	80.8%
Titles with one use only as % of total transactions	56.6%	44.3%	59.8%

Table 4.20 In-house survey results for one use only titles  
as a percentage of total titles and total  
transactions.

#### 4.7 Open Reserve Room Results

Data collected in this area was recorded with a view to comparison with external circulation data for the same period.

The reserve room caters for potentially high use titles. The usual sources for selections for reserve are reading lists supplied to students and titles that library staff recognise as being in high demand through external circulation.

Reserve room material is assumed to have a high probability of being intensively used.

Although the survey was limited, some comparison of the two sets of data was considered desirable to establish whether high use occurs concurrently from the two collections.

The survey results for the reserve section did not support the notion of high use.

The overall results for a three week survey were:

Discipline	Multiple Use Titles	Transactions For Multiple Use Titles	Titles Used Once	Total Transactions	Multiple Use Titles as % of Total Transactions	Single Use Titles as % of Total Transactions
330-339	328	1045	1165	2210	47.3	52.7
657-658	67	214	172	386	55.4	44.5
620-629	24	56	213	269	20.8	79.2
TOTAL	419	1315	1550	2865	46.0	54.0

TABLE 4.21 Multiple and single use on a title basis in Reserve Room for Economics/Management and Engineering for the period 30th September to 21st October, 1978.

In analysing the multiple-use statistics for each of the disciplines, it was found that the bulk of the multiple-use was between two and five transactions per title. In the economics discipline (330's) use ran as high as 17 loans for a title. However, between 5-17 uses there were very few titles involved (in fact one per cent of active titles).

In the management discipline (657-658) where the percentage of multiple use was higher than single use overall, the major proportion of this multiple use, 25 per cent, was for titles with two uses only. In the management area use ran as high as ten loans for a title. However, as for economics, between 5-10 uses, there were very few titles involved (two per cent of the active titles).

In engineering, the maximum use of a title was five, and titles with two uses accounted for 13 per cent of total use (multiple use was 21 per cent of the total).

The survey, although limited to a three week period, revealed no support for reserve room material being potentially high movers. This could have significant implications for the stocking of such collections if this effect applies generally. Reserve room collections often maintain more than one copy to deal with heavy usage on titles. An examination of the results of this survey raises the question of whether there would be any

advantage in removing copies from external circulation. Since the major proportion is single transaction use, and as material can only be borrowed for either four hours or overnight, there is no support here for the premise that the material is being used either extensively or for in-depth research.

One explanation that could be put forward is that material in reserve is known to be available at all times and, consequently, its major use is either quick reference or for photocopying relevant sections of material which appears on reading lists. (For the particular reserve room studied the latter use would show up in the records obtained).

If the majority of use is one or two occurrences, reserve sections may not be necessary as such low occurrences could easily be maintained by copies available for loan. The trade off would be in the salary savings obtained from the reserve room versus the extra copies that would be purchased where there is proven need.

A longer term analysis may reveal a different emphasis but for the purposes of this research the survey results did not warrant a comparison with circulation data. As the major portion of use was single occurrence, a comparison with high moving external circulation titles would not provide any information that would add to the development

of the concept of 'core'.

It is significant that reserve room collections are often, by default, considered to be the 'core' undergraduate collection because of supposed high use; the above results could place such a concept in dispute. Further, it cannot be assumed that reading list material need necessarily supply a 'core' in an undergraduate collection.

#### 4.8 References

1. Slote, R. Weeding Library Collections - II.  
Littleton, Colorado, Libraries Unlimited Inc., 1982,  
p.25, 26.
2. Grieder, Elmer M. The effect of book storage  
circulation services. College & Research Libraries  
v.2, October 1950, p.374-376.
3. The limitation of this result needs to be appreciated.  
First it does not imply that the circulation would be  
reduced by 3 or 4 per cent if the items were not  
available in open access. In many cases the copy  
chosen was one of a number which may have been  
regarded as comparable in value by the user.  
Accordingly, there would have been a substitution  
operation. Alternatively, if the specific work was  
sought for the historical insights that it may  
provide, then there would have been a request for  
retrieval from the closed access collection. The  
utility loss from this is difficult to estimate. It  
is necessary also to recall that there are benefits in  
removal of items with a low probability use. In  
summary therefore, the Grieder figures, though  
interesting, require interpretation.

4. Buckland, M. Book Availability and Library Use.  
N.Y., Pergamon Press, 1975, p.21
5. Ibid., p.27
6. Ibid., p.89-108.
7. Slote, op. cit., p.33. Slote's phrase "would satisfy" should not be misinterpreted. It is important to note that the reference is to current activity level and not to the activity that may result if access to certain parts of the collection was changed. If, for example, there was a program of weeding or item relocation, then it could be that the items which accounted for 70 per cent of circulation before the change, could satisfy 90 per cent of loans after the change.
8. Brookes, B.C. The Growth, Utility and Obsolescence of Scientific Periodical Literature. Journal of Documentation, Vol 24 (No. 4), Dec. 1970, p.283-294.
9. Slote, op. cit., p.43.

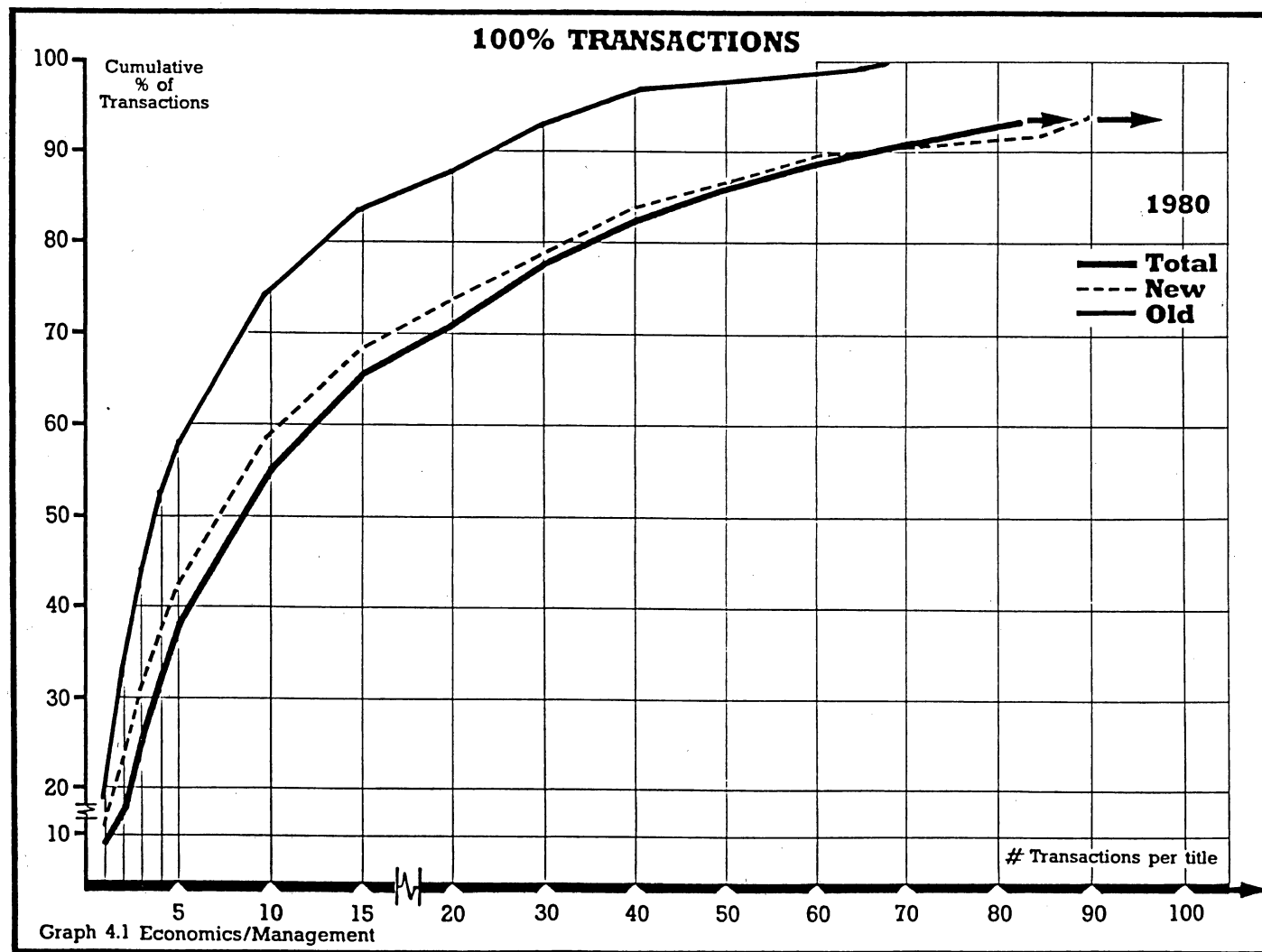
10. Undergraduates were able to be divided in this way as identification numbers made it possible to determine the year of enrolment. It is recognised that there may be admissions with advanced standing and there may be some users in the borrower classification. Nonetheless, the percentage of such errors was considered to be acceptable for the types of analyses of interest here.
11. Zero uses here indicate all titles that were active in 1980 but which were not used by the user group in question.
12. This thesis does not advocate the use of statistical procedures for decision-making. Rather, these techniques are seen as having a filtering role where attention is able to be confined to an appropriately small number of titles where the return from the investment of time in making a decision on those titles is warranted. "Everything" may not be a cost-effective answer.
13. Slote op. cit., p.25.
14. A weakness of this measure is that it concerns titles rather than transactions. This point is merely to be noted here and will not be explored in detail in this thesis.

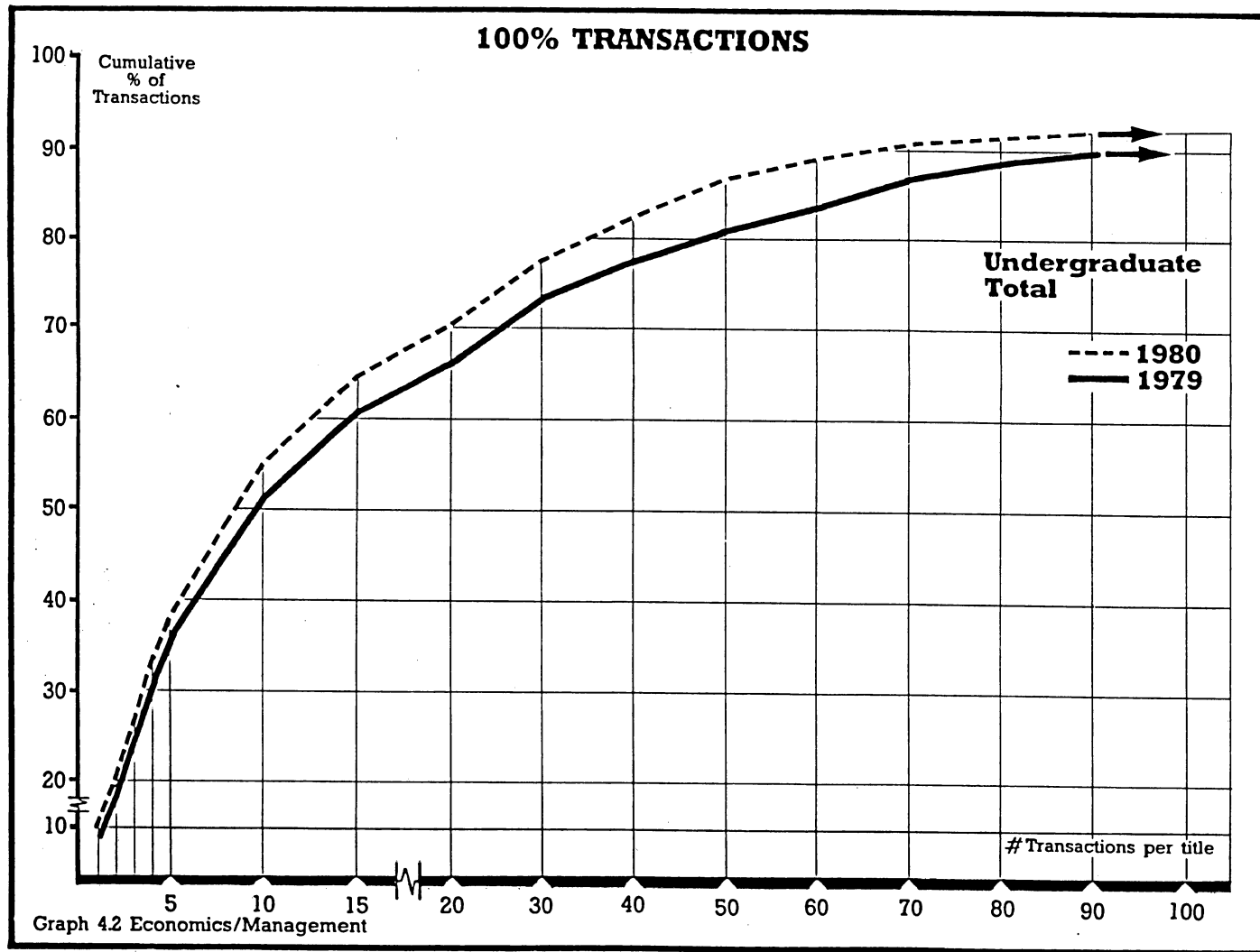
15. "Warehouse weed" is used here to describe the weeding of collections which have multiple copies of the titles held.
16. A lesser figure was taken for engineering titles because of the smaller holdings of titles that could be drawn upon. Even so, this figure gave roughly the same percentage figure as Economics/Management.
17. Use of Library Materials. The University of Pittsburg Study. N.Y., Dekker, 1979, p.49.
18. Ibid., p.49.
19. Morse, Philip. Library Effectiveness: A Systems Approach. Mass., M.I.T., 1968, p.9.

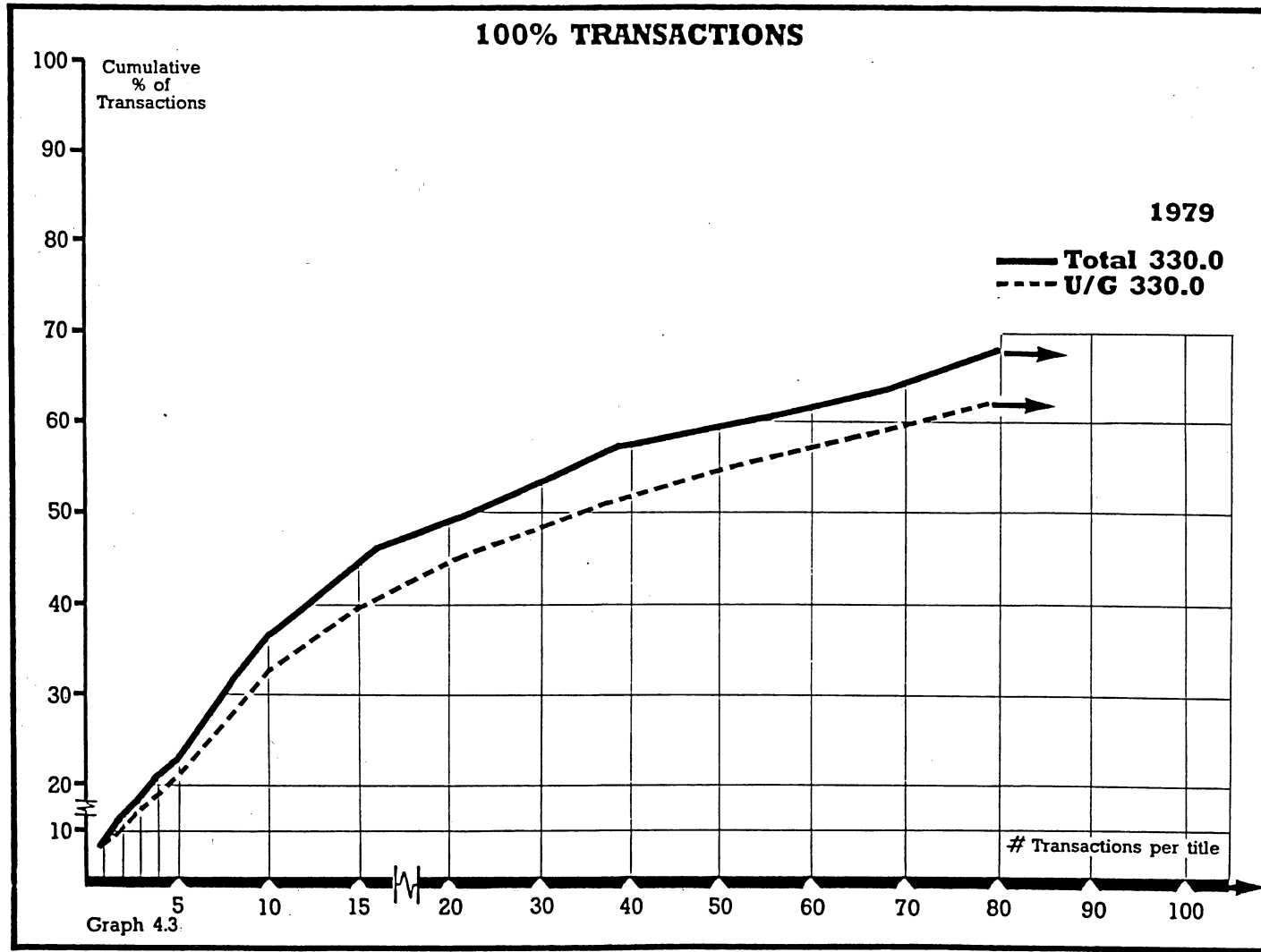
GRAPHS 4.1 to 4.7

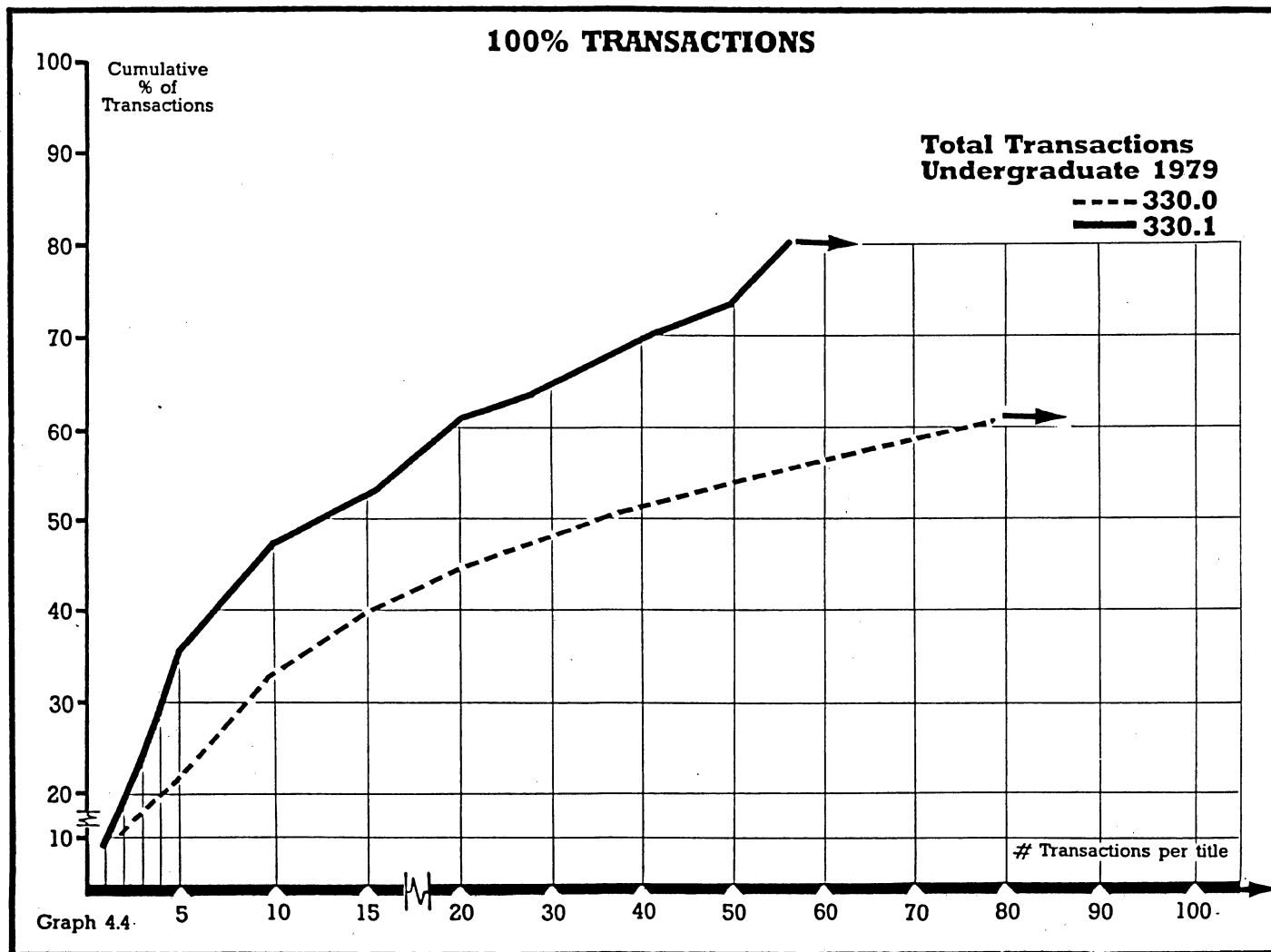
Note: The curves given overleaf were drawn through selected points on the relationships.

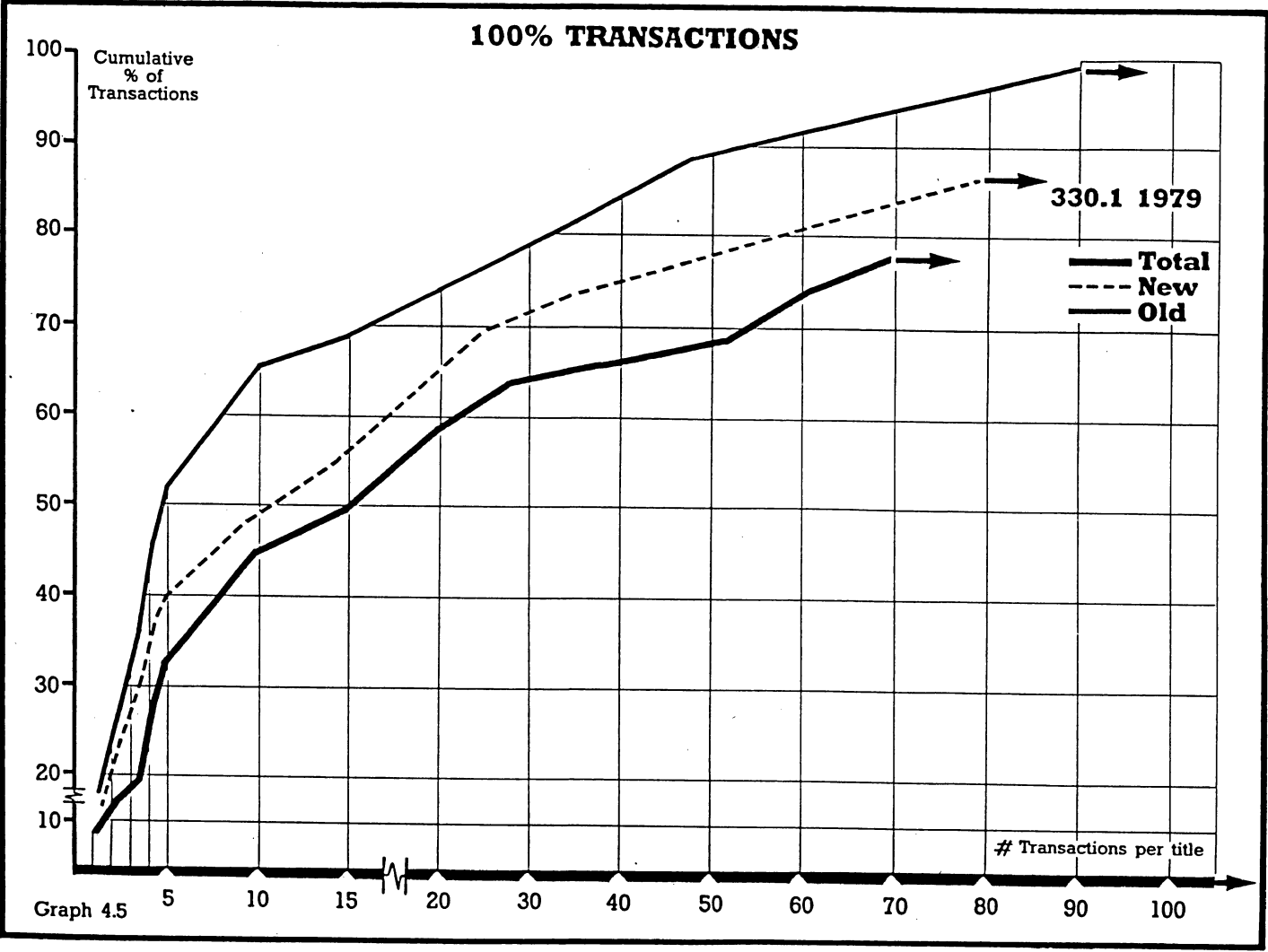
In all cases the relationship is only defined at integral values of the ordinate (transaction/title) and it should not be construed that there is a smooth relationship between the variables.

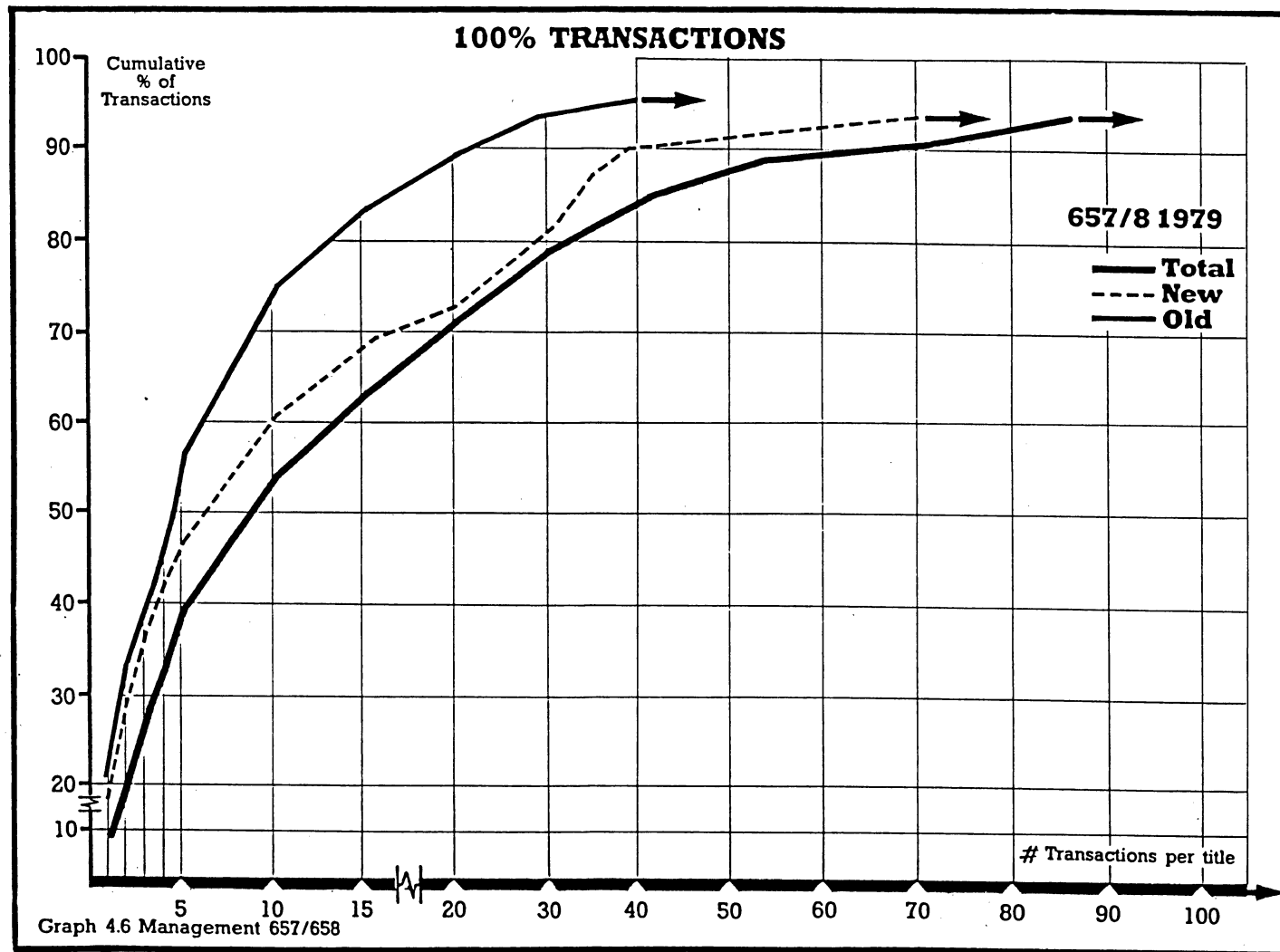


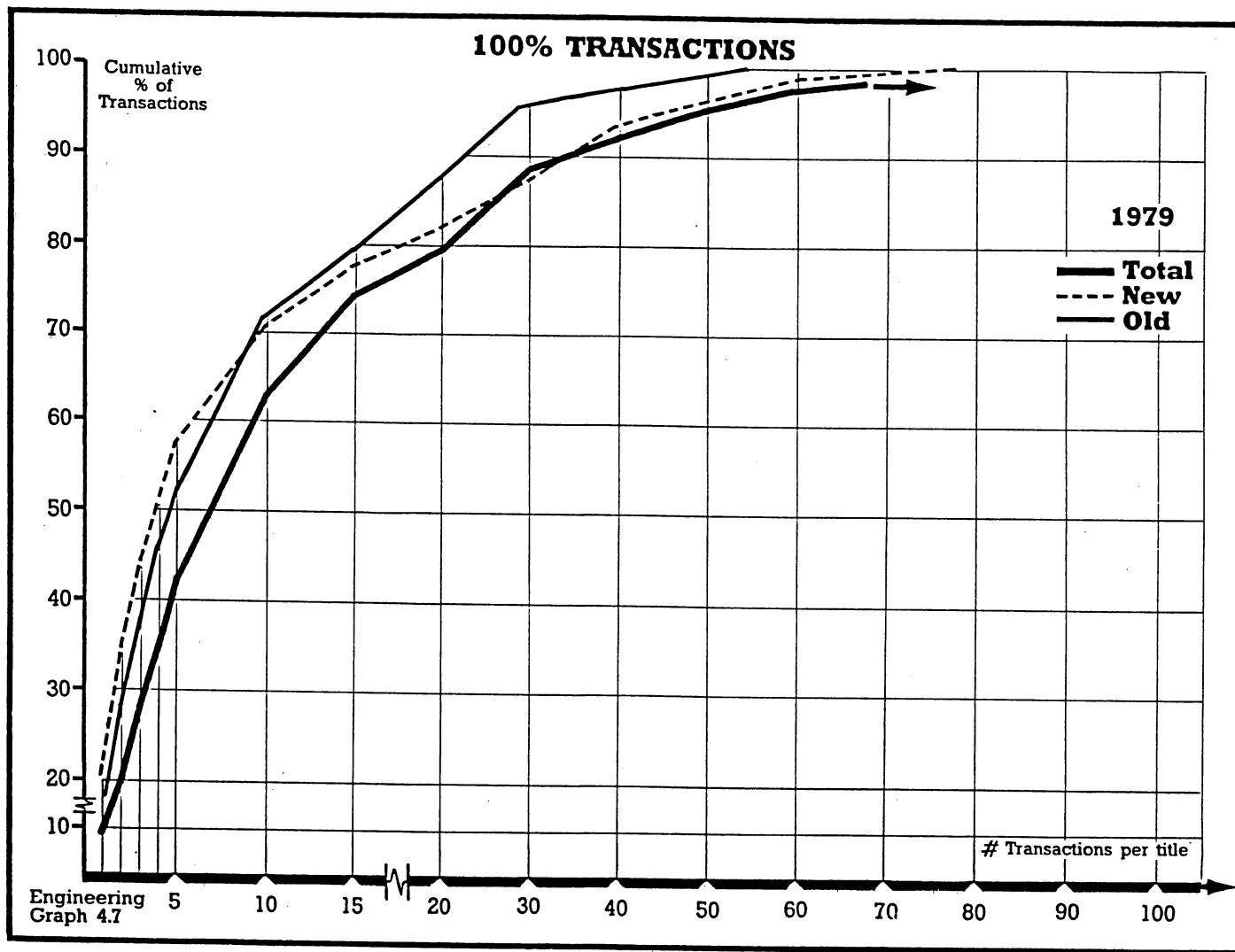












## APPENDIX 4.1

Distribution for 1979 Undergraduate Use

DISTRIBUTION FOR 1979 ACTIVITY

ATTRIBUTE 0679

TRANS =	0	# TITLES =	2018	CUMUL TRANS =	0	CUMUL TITLES =	2018
TRANS =	1	# TITLES =	3100	CUMUL TRANS =	3100	CUMUL TITLES =	5118
TRANS =	2	# TITLES =	1391	CUMUL TRANS =	5882	CUMUL TITLES =	6509
TRANS =	3	# TITLES =	707	CUMUL TRANS =	8003	CUMUL TITLES =	7216
TRANS =	4	# TITLES =	424	CUMUL TRANS =	9699	CUMUL TITLES =	7640
TRANS =	5	# TITLES =	317	CUMUL TRANS =	11284	CUMUL TITLES =	7957
TRANS =	6	# TITLES =	212	CUMUL TRANS =	12556	CUMUL TITLES =	8169
TRANS =	7	# TITLES =	159	CUMUL TRANS =	13669	CUMUL TITLES =	8328
TRANS =	8	# TITLES =	99	CUMUL TRANS =	14461	CUMUL TITLES =	8427
TRANS =	9	# TITLES =	80	CUMUL TRANS =	15181	CUMUL TITLES =	8507
TRANS =	10	# TITLES =	72	CUMUL TRANS =	15901	CUMUL TITLES =	8579
TRANS =	11	# TITLES =	71	CUMUL TRANS =	16682	CUMUL TITLES =	8650
TRANS =	12	# TITLES =	54	CUMUL TRANS =	17330	CUMUL TITLES =	8704
TRANS =	13	# TITLES =	46	CUMUL TRANS =	17928	CUMUL TITLES =	8750
TRANS =	14	# TITLES =	33	CUMUL TRANS =	18390	CUMUL TITLES =	8783
TRANS =	15	# TITLES =	28	CUMUL TRANS =	18810	CUMUL TITLES =	8811
TRANS =	16	# TITLES =	32	CUMUL TRANS =	19322	CUMUL TITLES =	8843
TRANS =	17	# TITLES =	24	CUMUL TRANS =	19730	CUMUL TITLES =	8867
TRANS =	18	# TITLES =	13	CUMUL TRANS =	19964	CUMUL TITLES =	8880
TRANS =	19	# TITLES =	15	CUMUL TRANS =	20249	CUMUL TITLES =	8895
TRANS =	20	# TITLES =	16	CUMUL TRANS =	20569	CUMUL TITLES =	8911
TRANS =	21	# TITLES =	13	CUMUL TRANS =	20842	CUMUL TITLES =	8924
TRANS =	22	# TITLES =	19	CUMUL TRANS =	21260	CUMUL TITLES =	8943
TRANS =	23	# TITLES =	9	CUMUL TRANS =	21467	CUMUL TITLES =	8952
TRANS =	24	# TITLES =	5	CUMUL TRANS =	21587	CUMUL TITLES =	8957
TRANS =	25	# TITLES =	9	CUMUL TRANS =	21812	CUMUL TITLES =	8966
TRANS =	26	# TITLES =	4	CUMUL TRANS =	21916	CUMUL TITLES =	8970
TRANS =	27	# TITLES =	11	CUMUL TRANS =	22213	CUMUL TITLES =	8981
TRANS =	28	# TITLES =	10	CUMUL TRANS =	22493	CUMUL TITLES =	8991
TRANS =	29	# TITLES =	6	CUMUL TRANS =	22667	CUMUL TITLES =	8997
TRANS =	30	# TITLES =	7	CUMUL TRANS =	22877	CUMUL TITLES =	9004
TRANS =	31	# TITLES =	5	CUMUL TRANS =	23032	CUMUL TITLES =	9009
TRANS =	32	# TITLES =	7	CUMUL TRANS =	23256	CUMUL TITLES =	9016
TRANS =	33	# TITLES =	6	CUMUL TRANS =	23454	CUMUL TITLES =	9022
TRANS =	34	# TITLES =	4	CUMUL TRANS =	23590	CUMUL TITLES =	9026
TRANS =	35	# TITLES =	3	CUMUL TRANS =	23695	CUMUL TITLES =	9029
TRANS =	36	# TITLES =	1	CUMUL TRANS =	23731	CUMUL TITLES =	9030
TRANS =	37	# TITLES =	4	CUMUL TRANS =	23879	CUMUL TITLES =	9034
TRANS =	38	# TITLES =	3	CUMUL TRANS =	23993	CUMUL TITLES =	9037

DISTRIBUTION FOR 1979 ACTIVITY

ATTRIBUTE U679

TRANS = 39	# TITLES = 3	CUMUL TRANS = 24110	CUMUL TITLES = 9040
TRANS = 40	# TITLES = 3	CUMUL TRANS = 24230	CUMUL TITLES = 9043
TRANS = 41	# TITLES = 1	CUMUL TRANS = 24271	CUMUL TITLES = 9044
TRANS = 42	# TITLES = 3	CUMUL TRANS = 24397	CUMUL TITLES = 9047
TRANS = 43	# TITLES = 1	CUMUL TRANS = 24440	CUMUL TITLES = 9048
TRANS = 44	# TITLES = 4	CUMUL TRANS = 24616	CUMUL TITLES = 9052
TRANS = 45	# TITLES = 3	CUMUL TRANS = 24751	CUMUL TITLES = 9055
TRANS = 46	# TITLES = 3	CUMUL TRANS = 24889	CUMUL TITLES = 9058
TRANS = 47	# TITLES = 3	CUMUL TRANS = 25030	CUMUL TITLES = 9061
TRANS = 48	# TITLES = 1	CUMUL TRANS = 25078	CUMUL TITLES = 9062
TRANS = 49	# TITLES = 1	CUMUL TRANS = 25127	CUMUL TITLES = 9063
TRANS = 50	# TITLES = 1	CUMUL TRANS = 25177	CUMUL TITLES = 9064
TRANS = 51	# TITLES = 1	CUMUL TRANS = 25228	CUMUL TITLES = 9065
TRANS = 52	# TITLES = 2	CUMUL TRANS = 25332	CUMUL TITLES = 9067
TRANS = 53	# TITLES = 2	CUMUL TRANS = 25438	CUMUL TITLES = 9069
TRANS = 54	# TITLES = 2	CUMUL TRANS = 25546	CUMUL TITLES = 9071
TRANS = 55	# TITLES = 3	CUMUL TRANS = 25711	CUMUL TITLES = 9074
TRANS = 56	# TITLES = 2	CUMUL TRANS = 25823	CUMUL TITLES = 9076
TRANS = 59	# TITLES = 1	CUMUL TRANS = 25882	CUMUL TITLES = 9077
TRANS = 61	# TITLES = 2	CUMUL TRANS = 26004	CUMUL TITLES = 9079
TRANS = 62	# TITLES = 2	CUMUL TRANS = 26128	CUMUL TITLES = 9081
TRANS = 63	# TITLES = 2	CUMUL TRANS = 26254	CUMUL TITLES = 9083
TRANS = 64	# TITLES = 1	CUMUL TRANS = 26318	CUMUL TITLES = 9084
TRANS = 65	# TITLES = 1	CUMUL TRANS = 26383	CUMUL TITLES = 9085
TRANS = 66	# TITLES = 2	CUMUL TRANS = 26515	CUMUL TITLES = 9087
TRANS = 67	# TITLES = 1	CUMUL TRANS = 26582	CUMUL TITLES = 9088
TRANS = 68	# TITLES = 1	CUMUL TRANS = 26650	CUMUL TITLES = 9089
TRANS = 69	# TITLES = 3	CUMUL TRANS = 26857	CUMUL TITLES = 9092
TRANS = 70	# TITLES = 1	CUMUL TRANS = 26927	CUMUL TITLES = 9093
TRANS = 71	# TITLES = 1	CUMUL TRANS = 26998	CUMUL TITLES = 9094
TRANS = 74	# TITLES = 3	CUMUL TRANS = 27220	CUMUL TITLES = 9097
TRANS = 75	# TITLES = 2	CUMUL TRANS = 27370	CUMUL TITLES = 9099
TRANS = 78	# TITLES = 1	CUMUL TRANS = 27448	CUMUL TITLES = 9100
TRANS = 79	# TITLES = 3	CUMUL TRANS = 27685	CUMUL TITLES = 9103
TRANS = 82	# TITLES = 1	CUMUL TRANS = 27767	CUMUL TITLES = 9104
TRANS = 89	# TITLES = 1	CUMUL TRANS = 27856	CUMUL TITLES = 9105
TRANS = 91	# TITLES = 1	CUMUL TRANS = 27947	CUMUL TITLES = 9106
TRANS = 93	# TITLES = 1	CUMUL TRANS = 28040	CUMUL TITLES = 9107
TRANS = 95	# TITLES = 1	CUMUL TRANS = 28135	CUMUL TITLES = 9108
TRANS = 96	# TITLES = 1	CUMUL TRANS = 28231	CUMUL TITLES = 9109

# DISTRIBUTION FOR 1979 ACTIVITY

## ATTRIBUTE 0679

TRANS = 98	# TITLES = 1	CUMUL TRANS = 28329	CUMUL TITLES = 9110
TRANS = 102	# TITLES = 1	CUMUL TRANS = 28431	CUMUL TITLES = 9111
TRANS = 108	# TITLES = 1	CUMUL TRANS = 28539	CUMUL TITLES = 9112
TRANS = 114	# TITLES = 1	CUMUL TRANS = 28653	CUMUL TITLES = 9113
TRANS = 116	# TITLES = 2	CUMUL TRANS = 28885	CUMUL TITLES = 9115
TRANS = 117	# TITLES = 1	CUMUL TRANS = 29002	CUMUL TITLES = 9116
TRANS = 127	# TITLES = 1	CUMUL TRANS = 29129	CUMUL TITLES = 9117
TRANS = 138	# TITLES = 1	CUMUL TRANS = 29267	CUMUL TITLES = 9118
TRANS = 140	# TITLES = 1	CUMUL TRANS = 29407	CUMUL TITLES = 9119
TRANS = 144	# TITLES = 1	CUMUL TRANS = 29551	CUMUL TITLES = 9120
TRANS = 152	# TITLES = 1	CUMUL TRANS = 29703	CUMUL TITLES = 9121
TRANS = 173	# TITLES = 1	CUMUL TRANS = 29876	CUMUL TITLES = 9122
TRANS = 203	# TITLES = 1	CUMUL TRANS = 30079	CUMUL TITLES = 9123
TRANS = 218	# TITLES = 1	CUMUL TRANS = 30297	CUMUL TITLES = 9124
TRANS = 231	# TITLES = 1	CUMUL TRANS = 30528	CUMUL TITLES = 9125
TRANS = 572	# TITLES = 1	CUMUL TRANS = 31100	CUMUL TITLES = 9126

OF RECORDS READ = 9126

AIL

## APPENDIX 4.2

### Terms Used

```
print thresh80
/*                                                                    */
/*      Create  told, tnew, total                                    */
/*                                                                    */
destroy told
range of y is year80
range of t is thresh80
retrieve into told (t.case,y.class,y.cutter,y.ctermin,y.no80)
where y.no80 > t.no
destroy tnew
retrieve into tnew (t.case,y.class,y.cutter,y.ctermin,y.nn80)
where y.nn80 > t.nn
destroy total
retrieve into total (t.case,y.class,y.cutter,y.ctermin,y.ug80)
where y.ug80 > t.ug
/*                                                                    */
/*      Find the items common to various groups                    */
/*                                                                    */
/*      Note : The following used restrictions that                */
/*      are equivalent to the joins.                                */
/*                                                                    */
destroy ttto
destroy ttnn
destroy ttcml
destroy ttcml2
/*      ttto -- intersection of total & old                        */
/*                                                                    */
retrieve into ttto(t.case, y.class,y.cutter,y.ctermin)
where y.no80 > t.no
and y.ug80 > t.ug
/*      ttnn -- intersection of total and new                      */
/*                                                                    */
retrieve into ttnn(t.case,y.class,y.cutter,y.ctermin)
where y.nn80 > t.nn
and y.ug80 > t.ug
/*      ttcml -- total & old U total & new                          */
/*      note that there will be some dups here                    */
/*                                                                    */
/*      For many calculations ttcml does not have to              */
/*      be determined. If required it can be found                */
/*      from the union operation. The number of                    */
/*      entries in ttcml is the sum of the number in              */
/*      ttto and ttnn.                                              */
/*                                                                    */
/*      ttcml2 -- the overlap between old & new                   */
/*                                                                    */
retrieve into ttcml2(t.case,y.class,y.cutter,y.ctermin)
where y.no80 > t.no
and y.nn80 > t.nn
```

```

/*      ttc2 -- the set common to o,n, total      */
range of t is thresh80
retrieve into ttc2(t.case,y.class,y.cutter,y.ctrm)
where y.no80 > t.no
and    y.nn80 > t.nn
and    y.ug80 > t.ug
/*
/*      Find the number of entries in told, tnew, total */
/*
destroy cnt
range of o is told
range of n is tnew
range of t is total
retrieve into cnto(o.case,ntold=count(o.class by o.case))
retrieve into cntn(n.case,ntnew=count(n.class by n.case))
retrieve into cntt(t.case,ntotal=count(t.class by t.case))
range of o is cnto
range of n is cntn
range of t is cntt
range of h is thresh80
retrieve into cnt(o.case,h.target,o.ntold,n.ntnew,t.ntotal)
where o.case = n.case
and    o.case = t.case
and    o.case = h.case
print cnt
destroy cnto
destroy cntn
destroy cntt
/*
/*      Find the number of entries in ttto, ttn, ttc1 */
/*      and ttc2. */
/*
destroy cntl
range of o is ttto
range of n is ttn
range of t is ttc1
range of s is ttc2
retrieve into cntlo(o.case,nttto=count(o.class by o.case))
retrieve into cntln(n.case,nttn=count(n.class by n.case))
retrieve into cntlt(t.case,nttc1=count(t.class by t.case))
retrieve into cntls(s.case,nttc2=count(s.class by s.case))
range of o is cntlo
range of n is cntln
range of t is cntlt
range of s is cntls
range of h is thresh80
retrieve into cntl(o.case,h.target,o.nttto,n.nttn,nttc=o.nttto + n.nttn,
    t.nttc1,s.nttc2)
where o.case = n.case
and    o.case = t.case
and    o.case = s.case
and    o.case = h.case
print cntl

```

```
destroy cntlo
destroy cntln
destroy cntlt
destroy cntls
destroy parameters
range of c is cnt
range of d is cntl
range of t is thresh80
retrieve into parameters(t.year,t.class,c.case,c.target, \
    overlap=float4(d.nttcml)/float4(c.ntold + c.ntnew - d.nttcml), \
    relevance=float4(d.nttcm - d.nttcm2)/float4(c.ntotal), \
    efficiency=float4(d.nttcm - d.nttcm2)/float4(c.ntold+cntnew-d.ntcml))
where c.case = d.case
print cnt
print cntl
print thresh80
print parameters
```

CHAPTER 5: CONCLUSION

		<u>Page</u>
5.1	Introduction.	284
5.2	Conclusions.	285
	5.2.1 Short-term Analysis.	285
	5.2.1a Distribution.	285
	5.2.1b Total or Group Analysis.	286
	5.2.2 Varying Threshold Approach.	287
	5.2.3 Group Characteristic Aproach.	288
	5.2.4 Long-term Analysis.	291
	5.2.5 Reserve and In-house use.	292
5.3	Data Limitations and Future Research.	294
	5.3.1 In-house Survey.	301
	5.3.2 Reserve Book Room Survey.	304
	5.3.3. In Brief.	306

## 5.1 Introduction

This thesis has primarily focused on the concept of 'core'. Through the analysis of circulation records for both short and long periods, it has looked at the acceptability of approximations to the analysis of behaviour of individual user groups.

It has also looked at in-house and reserve room data to add to the general results of the external circulation use patterns.

The analysis of use patterns to determine what books are used has significant implications for acquisitions and collection development. The applications presented in this thesis have concerned weeding, relegation and the setting of loan periods.

The conclusions are discussed in Section 5.2. The limitations of the data and topics for future research are dealt with in Section 5.3, and this section includes some discussion of both the in-house and reserve book room surveys.

The topics for future research which are outlined in this chapter complement the discussion that has been provided in that they illustrate how this work could be extended.

## 5.2 Conclusions

This section brings together the major conclusions of Chapter 4.

### 5.2.1 Short-term Analysis

#### 5.2.1a Distribution

In the initial process of establishing 'core' on a distribution basis, a number of cumulative distributions of activity were generated.

For the disciplines as a whole this analysis revealed typical diminishing behaviour.

As shown in Table 4.2b, if 'core' was defined on a transaction percentage level of 95 per cent (the figure often given in the literature), all titles borrowed in the year would need to have been included.

The minimum transaction level that would be required before some active titles were not included in 'core' would be 90 per cent in economics/management.

Tables 4.3b and 4.4b indicate that this also applies for the class breakdown in economics/management with the

exception of two cases: 330.9 for 'total' undergraduates in 1980 and 1979.

The minimum transaction level that would be required before some active titles were not included is 97 per cent. All items borrowed would, on this approach, be included in the core.

In short, the diminishing activity behaviour encountered in this investigation shows that for the library studies the figures mentioned in the literature may not be usable. This does not, however, mean that the approach of the literature is inappropriate.

If the task is weeding, then putting all borrowed titles in 'core' is not a problem. The results of Table 4.2 lead to a study of the multiple copy holdings of inactive titles. The values in the table also indicate that the decision on what constitutes core is not very sensitive to variations in the threshold.

#### 5.2.1b Total or Group Analysis

Various transaction thresholds were established for analysis of 'old', 'new' and 'total' undergraduate use. Tables 4.5, 4.6 and 4.7 give the results for an analysis over one year.

The results reveal that 'new' thresholds are, in most cases, substantially greater than the 'old' thresholds and have similar results over the range of transaction activity figures. Most of the thresholds for the 80 per cent level are '1' which is the smallest possible value. The implication of this is that one approach to "warehouse weeding" for transaction thresholds above 80 per cent in a one-year analysis would not be practical. It could, however, still be valuable in a weeding study to identify titles that had not been used.

#### 5.2.2 Varying Threshold Approach

This deals with the results relating to relevance and efficiency for changing thresholds.

Relevance was high and in 1979/1980 no figure fell below 90 per cent. This implies that for every 100 titles in 'total' core (i.e. the titles where all undergraduate use exceeds the 'total' threshold) 90 will also be in the 'combined' core.

Results for the classes do not differ markedly from the discipline results. There are only four cases where relevance is less than 90 per cent and these are mainly associated with 80 and 40 per cent transaction cut-off points.

Accordingly, titles selected for follow-up analysis on the basis of 'total' activity will not contain a substantial proportion of titles that would not have been selected as a result of a more comprehensive analysis. Note, however, that the cases where this applied were those where there was marked consistency in what was sought for user group behaviour. If this does not apply, the relevance is unlikely to be high.

Efficiency results at the 80 per cent cut-off point in economics/management show that 25 per cent (i.e.  $1/0.8 - 1$ ) of the titles in combined core will not be retrieved from the total undergraduate core. Thus undergraduate core will provide most, though not all, titles that would be of interest. The class analysis revealed similar results to the discipline results. Results appear in Tables 4.9 and 4.10.

#### 5.2.3. Group Characteristic Approach

These results are based on the 'old' threshold being held at '1' and 'new' and 'total' were all to vary as indicated in Table 4.5. In general, the analysis of this section may be considered typical of the case where the characteristics sought for various user group behaviour are different.

The relevance factor, as applied to the disciplines, remained high and, in fact, was higher than for the Varying Threshold Approach. The Group Characteristic Approach will effectively reduce the threshold for 'old' and, in so doing, increase the size of the combined core. It is therefore likely that titles in 'total' not in the varying core will be in the Group Characteristic core. This result provides confidence in the use of 'total' as an indicator of titles that will be in 'combined core'.

Efficiency is significantly lower than in the Varying Threshold approach. This is to be expected since the pool of titles being considered in 'old' is much higher than that obtained with any threshold value of 'VTR'. Results appear in Table 4.11.

The Varying Transaction Threshold is designed to look at high and low usage titles with the intention of providing an indication of which titles need to be further studied.

In examining the relegation problem, the Group Characteristic Approach should be used, as it established a 'core' transaction level that effectively includes all material borrowed for any reason. Note, however, the difference between the relegation and relocation problem.

In order to divide the collection into high and low use material, it is necessary to look at the overlap in the use

of titles by the sub-groups under review. This research found it necessary to look at the overlap between cores of 'old' and 'new'. Table 4.12 gives results for Ratio 1 and compares it with the value for overlap (old/new).

The results indicate low overlap (with no result above 29 per cent) which makes the location task straightforward as there is only a small percentage of titles of interest to both groups.

Table 4.12 shows that in engineering, the overlap tends to reduce with the percentage of total transactions. This reflects the difference between the 'high-movers' of the two user groups. When the transaction threshold is low, attention is focused on these high movers; when it is high a larger title population is captured.

Uniformity in the overlap measure (as shown by the 'VTR' calculations for economics/management) could be mainly a reflection of subject interest.

Table 4.13 gives overlap results for the 'VTR' cases for the classes. It may be noted that overlap tends to be class dependant. There tends to be a constant relationship between the overlap measure and Ratio 1 within a class. The figures for the population size indicate that the values need to be treated with caution.

#### 5.2.4 Long-term Analysis

This analysis examined whether there were use patterns that may assist exercises such as culling.

The analysis looked at the falling off of activity after the maximum. It was important to determine whether the pattern was cyclical and whether activity is, as a consequence, likely to rise once the maximum is passed.

Such analysis would provide some indication of how long it is necessary to wait before culling, without removing material that may be required in later years.

As for the short-term analysis, a cumulative distribution was generated for both titles and transactions. It revealed typical diminishing return behaviour.

Having determined an appropriate transaction level and fall-off point, two ratios were established. One calculated the ratio of use in adjacent years; the other looked at intervals separated by one year. In both cases, the denominator was taken from the year of (local) maximum activity.

If weeding was undertaken on the basis of adjacent year ratios, 37 titles would have been examined. If the chosen

drop-off figure was to be maintained in the second year, only 23 titles would have met the criterion.

Amongst the high movers, only a small pool was retrieved for culling analysis. To obtain an adequate pool, the threshold would need to be dropped. The view in some publications, that fall-off occurs quickly, was not supported by the data. In most cases there was a gradual reduction. Note that there were limitations in the data that meant this approach could not be applied to the extent desirable. The approach is much more promising than the above (37, 23) figures may suggest.

#### 5.2.5 Reserve and In-house use

Both the surveys undertaken were over short-time periods and are, therefore, illustrations only. However, there are several points that should be made.

The development of a rudimentary cluster approach for in-house use, as discussed in 4.6.2, showed the expected predominately intermittent single copy use. More data is required if the conjectures associated with the in-house survey, as discussed in Section 4.6.4, are to be resolved.

The reserve survey revealed no support for the proposition that reserve room materials are predominately high movers. This could have significant implications for the stocking

of such collections if this effect applies generally. Reserve room collections usually maintain more than one copy to deal with heavy usage. An examination of the results raises the question of whether there would be any advantage in removing copies from external circulation.

A more detailed analysis would be necessary to examine this issue with confidence.

### 5.3 Data Limitations and Future Research

In addition to the problems discussed in Chapter 3 associated with the processing of data, there were other features of the data that imposed limitations on the type of analysis possible.

The growth of the collection was disregarded. None of the analyses here used acquisition or cataloguing data from the University of New South Wales records. The study was 'synchronous' i.e., there was no allowance made for the different size of the collection over the period of analysis.

It was assumed that all materials in the disciplines reviewed had been classified consistently and material was not scattered through the collection. An analysis based on narrow bands in class numbers was not used because of the changes that had occurred in the Dewey schedules over the period. For the areas of Dewey considered, there were no major changes in classification tables during the period of the study.

No allowance was made in the study of actual use of heavily used material in the reserve collections. It was assumed that duplicates were available for loan and where they were not, that material in close proximity in the classification would be used as substitutes.

There were also a number of user constraints that need to be mentioned.

'Outside' influences on use could not be covered. For example, a new emphasis in a faculty resulting in changed reading patterns could not be followed up effectively unless it was known that there had been such a change. Further, it was assumed that the library being studied was the only source of material for users and for the students enrolled in the University's programs. Use of other collections was not and could not be controlled.

No attempt was made to maintain a record of students who had tried to obtain material from the collection but had failed. Correspondingly, there was no way of knowing from the available data whether the items borrowed were the items sought or whether they were substitutes.

With these data limitations and those expressed in Chapter 3, a number of possible analyses could not be undertaken and resulted in some gaps in the research. (It should, however, be noted that it is not being argued that these gaps should be fitted in 'practical' analyses as distinct from research studies. That is the issue of the cost of an analysis that needs to be controlled).

The importance of text and reference material in the analysis could not be checked since there were no

machine-readable lists of this material and as the subject enrolment of the particular student borrower was not available to the analysis.

Edition/copy analyses were not examined as the transaction data did not contain edition numbers and there was no file accessible to the analysis that related copy number and edition number. Accordingly, questions such as the relationship between activity level and edition could not be probed.

Likewise, the relationship between the age of the title and its transaction level (as discussed in Kent<sub>1</sub>) could not be studied. (There was no file available to the analysis containing details such as the date received, the dates of the first and subsequent editions, and the relationship of this data to the copy number).

Any subsequent analysis of 'core' should consider 'aging' of titles and use patterns as discussed by Kent<sub>2</sub>.

Further, the role of reading lists and their continuous use over several years should be reviewed to establish whether such practices create distortions in the use of collections. (This may be an 'education' question rather than a 'librarianship' question). It would be an interesting research study to look at the relationship of text and reference book status on patterns encountered in a long-term analysis.

There was no analysis of students that distinguished between full and part time enrolments. The grouping of students into 'old' and 'new' categories was based on the year the student enrolled. Consequently there is no indication as to the level of study reached by part time students. Of course there is an assumption of some regular pattern of study that underlies this categorisation. It may have been the case that this was sustainable in the 1970s but that it cannot be strongly argued after this time.

Further research could examine specific user groups, e.g. students enrolled in first year Economics, to provide a more detailed analysis of particular group behaviour than the 'old' and 'new' undergraduate groups studied here. However, given that 'total' has been shown to be satisfactory for many purposes, then it may be conjectured that a study based on enrolment is unlikely to provide differing results. Again the issue of the cost of providing the additional detail in the analysis is one that needs to be pursued. Here too, access is needed to an enrolment or borrowers' file; questions of privacy are valid.

The important grouping required was undergraduate versus specialised collections. Hence, it was not considered necessary to look at the physical or social sciences collections separately. (In fact, for the two disciplines

studied there was largely only one research library holding). Further research may reveal different results but for the purposes of the theory developed here, distinct collection analysis was not undertaken. Note that there are two separate issues here. One concerns the acquisition of a title; the other is related to the location of the copies obtained.

Any future research could also look at use in specific periods, e.g. half semesters, exam periods and assignment periods. The present analysis looks at full year data and, in so doing, to some extent 'averages' out the results through heavy and light periods of academic use.

No distinction is drawn between the servicing of a demand that occurs twice in a year and one leading to the same total number of transactions, but which occurs at one time only.

An analysis of specific periods may well provide use patterns that could have an influence on collection development other than those periods dealt with here. In essence that would allow attention to be directed to the requirement that the number of copies held needs to be a function of the maximum demand and not the average demand.

Because of data deficiencies discussed in Chapter 3, the long term analysis did not have sufficient data to provide

more than an indication of how to proceed. Data for a period of, say, six consecutive years, is essential for this type of study. It provides more reliability regarding use cycles and it is apparent from the analysis in Chapter 4 that it offers significant advantages over an analysis based on a single year only.

It should be noted that the analyses looked at titles not transactions and this sometimes might be misleading.

For example, libraries weed collections for a number of reasons. If the intention is to look at titles and to maintain a collection which provides a diversity of thought, then an analysis using titles alone is appropriate.

If relevancy and efficiency measures are applied to the two cores separately or to 'total' on a title basis and a result of 70 per cent is obtained, this implies that three titles out of ten will be missed. This presents no problem since the loss in titles is not significant and the effect on the collection minimal; there would still be adequate titles to achieve the library's aim.

However, most libraries weed on a usage basis with the purpose of removing material that is not used. In this case an analysis by title would not be as apt as one which looks at transactions. If relevancy and efficiency figures

are applied to the same cores as discussed above but, in this case using transactions and the result is 97 per cent, there would be no need for concern. Although 30 per cent of the titles may have been missed in the title analysis, a 97 per cent transaction result indicates that the major part of the use would have been covered.

However, had this figure been 60 per cent then some action would have been necessary. A result such as this shows that not only would have 30 per cent of the titles been missed, but a major portion of use is also not being covered.

If the library's concern was to get usage correct and to check the efficiency of copies within the collection, then the title analysis would not have been adequate; it would have been important to look at transactions also.

It should be noted that the concepts of relevance and efficiency developed here are 'hard'. That is, a title is either correctly or incorrectly classified in terms of 'core'. For the purposes of the analysis undertaken here, this presented no problem. No weighting of errors was used but, depending on the application of such measures, misclassifications may present problems and any future research should not exclude the possibility of having to look at the nature of errors.

### 5.3.1. In-house Survey

The purpose for examining in-house use was to establish whether:

1. material used in-house displayed intensive use,
2. it reflected browsing once-off use,
3. there was movement between sub-collections in the University library.

One of the major problems encountered in the in-house survey was data control. Although each sub-collection has its own pick-up and distribution area for sorting and reshelving material used in-house, much of the material is physically removed from its home floor and used elsewhere in the building. This results in long time-lags between use and sorting and often it was found that the time delays caused material to be counted twice resulting in duplicated samples. Although not a major problem in this research because of the shortness of the period under review, future research would need tighter controls to avoid this sampling duplication. Note that account needs to be taken of the shelving interval in determining the appropriate definition of clusters.

The sampling interval is also critical. Although a three week period was considered adequate here, in extensions of this type of analysis such an interval may not be

appropriate. Ideally, future research would need to consider at least one full academic year. Such a time frame would permit studies of features not discussed in this thesis, e.g. borrowing prior to examinations, high and low periods of borrowing and their relationship to examinations, assignments and vacation. The cost-benefit of such a survey would need to be assessed. It could be that what is desirable from the view of supporting a research study could be too long to be affordable for a practical analysis.

Kent<sub>3</sub> has referred to the assumption that patrons do not reshelve material; any future research would need to place tight controls on this. No comparison was made in this research between items used in-house and those used externally.

From the rudimentary cluster analysis developed in Section 4.6, the results indicated high single copy intermittent use of titles. It was not necessary to extend this further here. Additional questions for future examination are: What items are being used in-house that have not circulated externally; and, is the result likely to change the overall results of external circulation?

A number of conjectures are made in Section 4.6.4. A more extensive sample would provide more conclusive results as to: Whether the use of material in-house is a

browsing exercise designed to make choices from the alternative titles available; whether the use across collections is a result of saturation in one collection; and whether patterns vary between the hard and soft sciences. Future research regarding in-house circulation needs to look at these conjectures.

The in-house analysis in Section 4.6.2 includes a rudimentary cluster analysis. This was designed to examine alternative combinations of use and to place titles into specified categories of use.

The cluster patterns developed distinguished between intermittent and concentrated use and cases between these two extremes. For the purposes of this thesis it was not considered necessary to define or extend the discussion to the mathematical detail of cluster analysis.

Any future research into in-house studies would need first to refine the cluster analysis. A more detailed analysis using such a theory would require additional data and need to overcome the problems of data collection experienced here.

### 5.3.2 Reserve Book Room Survey

A review of reserve book room use was undertaken to look at the justification for placing material in reserve. Most material is placed in reserve on the premise that it will be high use.

The Kent<sub>4</sub> study looked at circulation book use patterns in a partial way. It presented results for circulation patterns for charges from the collection to the reserve Book Room but was not able to look at use in reserve by patrons.

The data available in this thesis was not so limited and Section 4.6 looked at internal circulation of the reserve Book material.

The limitation was that the data was not recorded in any automated format and required manual capture.

The labour cost of such a task allowed only a three week analysis. The results in Section 4.6 show that the use did not display high activity for many titles. The period chosen was expected to be a likely 'high use' time in the semester. (It should be noted that second semester use figures were lower in the late 1970s than corresponding first semester loans.)

Any future study of reserve room collections would ideally need to look at least at a full semester.

If of interest, it would then be possible to compare external borrowing and reserve use on a title or class basis. Note that it would not be possible to determine who the borrowers of reserve material were, though some reasonable conjectures could be made.

As for the full year analysis of external circulation, there would be a great advantage in looking at high peak and low use periods separately to establish whether use patterns differ significantly. The results in Section 4.6 indicate high single transaction use: Only a full semester period analysis could establish whether the period reviewed was atypical.

Another problem that would need to be examined is the inability to establish from manual records whether the loans are renewals. This could be significant. Loans were only for a four hour period. If in some cases the loans are renewals by one person for one or more periods this could have further implications for the stocking and establishing of reserve collections. (Such a study could address conjectures on the inherent difference in requirements for use of material from reserve and from external borrowing. In this view the reserve copies can

be heavily used when there are no copies available from the borrowable collection.)

To deal with the capture of this type of data efficiently it would be more cost-effective to automate the procedure than to use manual recording procedures.

### 5.3.3 In Brief

In general there are two main avenues of development of the areas dealt with in this thesis.

It is possible to look in more detail at a particular type of use (e.g. external borrowing, in-house or from reserve) and to examine typical library problems for that use.

For example, in the case of external borrowing there are issues such as weeding, collection development, loan period policies and the like. Here one needs to note that the standard problems differ between the different types of use.

One can also examine the interrelationship of the uses. One can probe the extent to which the user's behaviour reflects his initial aim for entering the library or whether it is a second-best solution to his needs.

The statistical approach of this thesis is (with some development) suited to the first class of development. Different analytical vehicles may be required in the second case.

In both cases there is need to look at cost-benefit trade offs. Collecting data is an expensive process. Following up massive amounts of data from some algorithm is probably even more costly. In addition to the obvious costs of these exercises, there is also the difficulty in quantifying effects on the users.

It is not difficult to establish theory. The real challenge may lie in including cost-effectiveness.

ADDITIONAL READING

The following are additional references which, although not cited, assisted as background reading in the development of the theory presented in this thesis.

1. AVRAMESCU, A.            Actuality and Obsolescence of Scientific Literature. J.A.S.I.S., v. 30(5) 1979, p. 296-303.
2. BOWER, C.A.            Patterns of use of the serial literature at the BLLD. BLL Review, v. 4(2) 1976, p. 31-36.
3. BROOKES, B.C.            The growth, utility and obsolescence of Scientific periodical literature. The Journal of Documentation, v. 26(4) Dec. 1970, p. 283-294.
4. BUCKLAND, M.K.            Are obsolescence and scattering related? Journal of Documentation, v. 28(3) 1972, p. 242.
5. BULICK, S. et.al.        Use of Library Materials in Terms of Age. J.A.S.I.S., v. 27 1976, p. 175-177.
6. BURRELL, Q.L.            The 80/20 Rule: Library Lore or Statistical Law? Journal of Documentation, v. 41(1) March 1985, p. 24-39.
7. CAWKELL, A.E.            Documentation note: Citations, obsolescence, enduring articles, multiple authorships. Journal of Documentation, v. 32(1) March 1976, p. 53-58.

8. CHEN, C.C. The use patterns of physics journals in a large academic research library. J.A.S.I.S., v. 23(4) July-Aug 1972, p. 254-265.
9. CLARK, C.V. Obsolescence of the patent literature. Journal of Documentation, v. 32 (1) March 1976, p. 32-52
10. FUSSLER, H. and SIMON, J.L. Patterns in the use of books in large research libraries, Chicago, University Press, 1969.
11. GALVIN, T.J. et.al. Use of a university library collection. A progress report on the Kent Study. Library Journal v. 102, Nov. 15 1977, p. 2317-2320.
12. GORE, D. Farewell to Alexandria: solutions to space, growth and performance problems of libraries. College and Research Libraries v. 37 1976, p. 556-558.
13. GRIEDER, E.M. The effect of book storage on circulation service. College and Research libraries v. 11 1950, p. 374-376.
14. GRIFFITH, B.C. et.al. The aging of scientific literature: a citation analysis. Journal of Documentation v. 35(3) Sept. 1979, p. 179-196.
15. HARRIS, C. A comparison of issues and in-library use of books. Aslib Proceedings, v. 29(3) March 1977, p. 118-126.
16. KENT, A. et.al. A commentary on "Report on the study of library use at Pitt by Professor Allen Kent et.al.," The Senate Library Committee, University of Pittsburg, July 1969. Library Acquisitions: Practice and Theory, v. 4 1980, p. 87-99.

17. KENT, A. A Rebuttal. Journal of Academic Librarianship, v. 7(12) Jan 1975, p. 12-30.
18. KOHUT, J.J. A comparative analysis of obsolescence patterns of the U.S. Geoscience literature. J.A.S.I.S. v. 25(4) July-August 1974, p. 242-251.
19. LAWANI, S.M. Obsolescence of American and French Agronomic Literatures. Journal of Librarianship, v. 7(1) Jan 1975, p. 12-30.
20. LINE, M.B. and CARTER, B. Changes in the use of sociological articles with time, a comparison of diachronous and synchronous data. BLL Review, v. 2(4) Oct 1974, p. 124-129.
21. LINE, M.B. and SANDERSON, A. Progress in documentation 'Obsolescence and changes in the literature with time'. Journal of Documentation, v. 30(3) Sept. 1974, p. 283-350.
22. LINE, M.B. and SANDERSON, A. Practical interpretation of citation and library use studies. College and Research Libraries, v. 35(5) Sept. 1975, p. 393.
23. McGRATH, W.E. Correlating the subjects of books taken out of and books used within an open-stack library. College and Research Libraries, v. 31 July 1971, p. 280-285.
24. McGRATH, W.E. Circulation studies and development: problems of methodology, theory and typology for reach. In Foundations in library and Information Studies, V. 10 Part B. JAI Press, 1980.
25. MARSHALL, P. How much, how often? College and Research Libraries, v. 35(6) 1975, p. 453-456.

26. MEADOWS, A.J. The citation characteristics of astronomical research literature. Journal of Documentation, v. 23 1967, p. 28-33.
27. METZ, P. The use of the general collection in the Library of Congress. Library Quarterly, v. 49(4) 1979, p. 415-434.
28. METZ, P. The Landscape of Literatures. Chicago, American Library Association, 1983.
29. O'NEILL, E.T. The effect of demand level on the optimal size of journal collections. Collection Management, v. 2(3) Fall 1978, p. 205-216.
30. RAO, R.I.K. Growth of periodicals and obsolescence of articles in periodicals: a case study in sociology. Library Science Slant document, v. 11(2) June 1974, p. 92-97.
31. RAO, R.I.K. Obsolescence and utility factors of periodical publications: a case study. Library Science Slant Document, v. 10(3) Sept. 1973, p. 297-307.
32. SANDISON, A. Densities of use and absence of obsolescence in physics journals at M.I.T. J.A.S.I.S., v. 25(3) May-June 1974, p. 172-182.
33. TRUESWELL, R.W. and TURNER, S.J. Simulating circulation-use characteristic curves using circulation data. J.A.S.I.S., v. 30 1979, p. 83-87.
34. TRUSWELL, R.W. Some circulation data from a research library. College and Research Libraries, v. 29 1968, p. 493-495.

35. TURNER, S.J. Trueswell's weeding technique: the facts. College and Research Libraries, v. 41(2) March 1980, p. 134-138.
36. URQUHART, J.A. and URQUHART, N.C. Relegation and Stock Control in Libraries, Boston, Oriel Press, 1976.
37. WHITE, G.T. Quantitative measures of library effectiveness. The Journal of Academic Librarianship, v. 3(3) 1977, p. 128-136.
38. WHITE, H.S. Library management in the tight budget seventies. Bulletin of the Medical Library Association, v. 65(1) Jan. 1977, p. 6-12.