

DOCUMENT RESUME

ED 071 674

LI 004 069

AUTHOR Clague, P.  
TITLE INSPEC SDI Investigation, 1967-1969. Volume III.  
INSTITUTION Institution of Electrical Engineers, London  
(England).  
SPONS AGENCY Office for Scientific and Technical Information,  
London (England).  
REPORT NO INSPEC-R-71-6  
PUB DATE 71  
NOTE 130p. (0 References); Chapters 17 to 31  
AVAILABLE FROM INSPEC, The Institution of Electrical Engineers,  
Savoy Place, London, WC2R 0BL, England (5 volumes, HC  
\$30.86)

EDRS PRICE MF-\$0.65 HC Not Available from EDRS.  
DESCRIPTORS Economics; Electronics; Foreign Countries;  
\*Information Dissemination; \*Information Services;  
\*International Programs; Physics; \*Relevance  
(Information Retrieval); \*Use Studies  
IDENTIFIERS England; INSPEC; Scientific and Technical  
Information; SDI; \*Selective Dissemination of  
Information

ABSTRACT

The third volume of this five volume study of the INSPEC SDI system contains: Recall performance of SDI profiles; Assessment of the Recall performance of the SDI system in relation to particularly valuable articles; Subjective assessment of the SDI system Recall performance; Relation between Precision performance of the system and document indexing; User assessment of the SDI service; Relationship between quality of user's initial statement of information requirements and the subsequent assessment of his profile; Effect of profile analysis and modifications on user's satisfaction with service; Effect of number of document notifications received by user on his relevance assessment; Comparison of relevance assessments based on the full document with those based on card notifications giving varying amounts of document information; Effect of including author names as search terms in SDI profiles; Variation of user satisfaction with compiler of profile; Methods of user interaction in profile compilation; Use made of SDI notifications; Desirability of including material other than English-language periodical articles in the coverage of the SDI service; and Novelty of information provided by the SDI service. (Volumes I and II are: LI004057 and 004068, Volume IV and V are: LI004070 and 004071.) (Author/NH)

Report No. R.70/6

INTERNATIONAL  
INFORMATION SERVICES  
IN PHYSICS,  
ELECTROTECHNOLOGY,  
COMPUTERS AND CONTROL

ED 071674

U.S. DEPARTMENT OF HEALTH,  
EDUCATION & WELFARE  
OFFICE OF EDUCATION  
THIS DOCUMENT HAS BEEN REPRO-  
DUCED EXACTLY AS RECEIVED FROM  
THE PERSON OR ORGANIZATION ORIG-  
INATING IT. POINTS OF VIEW OR OPIN-  
IONS STATED DO NOT NECESSARILY  
REPRESENT OFFICIAL OFFICE OF EDU-  
CATION POSITION OR POLICY

PERMISSION TO REPRODUCE THIS COPY-  
RIGHTED MATERIAL BY MICROFICHE ONLY  
HAS BEEN GRANTED BY  
Institution of  
Electrical Engineers  
TO ERIC AND ORGANIZATIONS OPERATING  
UNDER AGREEMENTS WITH THE US OFFICE  
OF EDUCATION. FURTHER REPRODUCTION  
OUTSIDE THE ERIC SYSTEM REQUIRES PER-  
MISSION OF THE COPYRIGHT OWNER

INSPEC SDI INVESTIGATION

1967 - 1969

Volume III

P Clague

The work reported in this documents was supported  
by a grant from the Office for Scientific and  
Technical Information (Ref.SI/20/15) of the  
Department of Education and Science.

© 1971 The Institution of Electrical Engineer.,  
Savoy Place, London WC2R 0BL, England.

ISBN 0 85296405 6

LJ 004 069

## TABLE OF CONTENTS

### Volume I

Preface

Chapter

page

1	Aims and background of the INSPEC SDI Investigation	1-1
2	General review	2-1
3	Selection of user and control groups	3-1
4	Loss of participants throughout Investigation	4-1
5	Reasons for loss of participants receiving SDI service, November 1968-November 1969	5-1
6	SDI system operation	6-1
7	Delays in service	7-1
8	Computer service for the SDI Investigation	8-1
9	Profile analysis and modification	9-1
10	Interaction with participants	10-1
11	User comments on the service	11-1

### Volume II

12	Composition of user and control groups and the information use habits of users and controls	12-1
13	Thesaurus development and document indexing	13-1

14	Problems of printing notification cards and assessing numbers required for each document	14-1
15	Recall and Precision failure	15-1
16	Precision performance of the SDI service	16-1

Volume III

17	Recall performance of SDI profiles	17-1
18	Assessment of the Recall performance of the SDI system in relation to particularly valuable articles	18-1
19	Subjective assessment of SDI system Recall performance	19-1
20	Relation between Precision performance of the system and document indexing	20-1
21	User assessment of the SDI service	21-1
22	Relationship between quality of user's initial statement of information requirements and the subsequent assessment of his profile	22-1
23	Effect of profile analysis and modifications on user's satisfaction with service	23-1
24	Effect of number of document notifications received by user on his relevance assessment	24-1
25	Comparison of relevance assessments based on the full document with those based on card notifications giving varying amounts of document information	25-1
26	Effect of including author names as search terms in SDI profiles	26-1
27	Variation of user satisfaction with compiler of profile	27-1
28	Methods of user interaction in profile compilation	28-1

29	Use made of SDI notifications	29-1
30	Desirability of including material other than English-language periodical articles in the coverage of the SDI service	30-1
31	Novelty of information provided by the SDI service	31-1

Appendix

Volume IV

1A	A proposal to investigate the selective dissemination of information	1A-1
3A	Covering letter to questionnaire	3A-1
3B	Questionnaire: survey of information use	3B-1
3C	Chasing letter	3C-1
3D	Letter of invitation to participate	3D-1
3E	Chasing letter	3E-1
3F	Statement of information requirements	3F-1
3G	Letter to heads of university departments	3G-1
3H	Electronics research workers in universities and technical colleges	3H-1
3I	Details of sampling	3I-1
3K	The role of the Project Associate	3K-1
4A	Size of sample: memorandum	4A-1
5A	Reasons for withdrawal	5A-1
6A	Periodical scanned for the service	6A-1
8A)	Sundry memoranda	8A-1
8B)		
8C)		
8D)		
9A	SDI Service questionnaire	9A-1
9B	Profile analysis and modification	9B-1

Volume V

10A	Communications with users etc.	10A-1
13A	Specimen sheets from descriptor file	13A-1
23A	Time on profile modification and analysis, and user assessment	23A-1
27A	Profiles compiled by compilers 1, 2 and 3.	27A-1
28A	Proposed comparison of methods of user interaction in compiling and modifying profiles	28A-1
28B	Letter inviting comments on draft profile	28B-1
28C	Explanatory letter re. above	28C-1
28D	Information Requirements - Queries	28D-1
28E	Notes for visits/interviews	28E-1
32A	Alternative form of SDI notifications	32A-1

## Chapter 17

### RECALL PERFORMANCE OF SDI PROFILES

Assessment of the percentage Recall performance of profiles in an SDI service is undoubtedly difficult and time-consuming, and more, perhaps, than any other figures derived from the system, the Recall figures are sensitive to the methods used to obtain them.

The general method used in the SDI Investigation was to send to users a list of a sample of the documents input to a given matching run, asking them to mark these documents which each would expect his profile to select. The marked documents were then compared with those which had been selected by his profile and thus the percentage Recall established.

There are several major drawbacks to this method:-

- 1) The number of documents relevant to the particular user and contained in the list is likely to be very small, resulting in large variations in the figures derived. At the same time the total number of items in the list may still be too large for the user to scan quickly and make consistent relevance assessments.
- 2) The composition of a sample list of documents based on a week's or a fortnight's input to the system will itself vary very considerably and give uneven results.

- 3) The user will seldom make the relevance assessment directly on the basis of his interests as stated in the profile but on a looser basis depending on which articles strike him as of interest at the time of the assessment.

Use of a large list of documents covering a larger time-span may, to some extent, meet problems 1 and 2 above but at the same time create others.

Only in special circumstances is it possible to ask users to check a large list item by item. In most cases the larger list needs to be classified to allow the user to find the small number of items on subjects of potential relevance to him. This in itself confines the possible choice of items to those in sections of the list which he can recognise from the classification system as likely to contain relevant articles. This being so it might be supposed that, depending on his particular interests and the type of classification used, his selection of items might be biased towards his chief, perhaps easily classified and therefore retrievable interests.

In the early stages of the Investigation the main purpose of the lists sent out to users was less to establish a figure for the Recall performance of the individual profile or the system as a whole, but more to discover, for the purpose of profile modification, what defects existed in the profiles. Owing to the fact that major testing and modification was still being done to profiles even after the start of the weekly service this continued to be the main use made of the lists for some two months after the start of the operational weekly service.

#### Subject lists

In addition to its index terms or descriptors, each document input to the system was assigned a set of up to three subject codes selected from the list in Figure 4.

It should perhaps be emphasized that the list of subject and the groupings shown are not to be considered as a strict classification. They were designed very quickly as a means of roughly sub-dividing the subject area covered by the investigation. To a large extent the groupings put together subjects which tended to be grouped together in user's profiles. However some of the groups are 'miscellaneous' categories in which a number of fairly unconnected subjects have been put for convenience simply to keep down the number of groups.

Using these codes for sorting, the computer was able to produce listings of the documents input to a given week's run, in the broad subject groups.

The interests of each user were similarly classified using the same broad subject classification codes. As many as four codes per user could be used for this purpose.

When it was wished to obtain Recall data for any particular week the 'Bulletin listing' was obtained from the computer on offset litho masters, and copies produced of lists for each particular classification code. Users were then sent copies of the lists for the appropriate subject codes with a covering letter asking them to mark relevant documents.

It may be imagined that the task of duplicating, selecting and despatching unique combinations of lists for each user was no small one, particularly when the total sample of users or a large proportion of them were involved. However this was accepted initially to obtain the necessary data for profile analysis and modification. The weeks for which this data was collected were 014 and 023 i.e. the fourth and thirteenth week of operational service since this period started at week 011.

For the 014 week lists the users were asked to mark relevant items but at week 023 they were asked to distinguish between the highly relevant (R1) and the partially relevant (R2) documents. The Recall figures derived were, for week 014, 52 percent and for week 023, 58 percent (Recall for R1 documents) and 45 percent (Recall of R1/R2 documents).

This tends to show that, as might be supposed, users mark more documents when they have a two-category marking system rather than a single category one, since borderline items which might be rejected in the latter system could be legitimately included as R2 in the former.

As soon as the major work of profile testing and modification had been completed it was decided to abandon the system of sending the subject Bulletin listings which required so much effort to produce, select and distribute. Instead it was decided to adopt a system of random listings in which a sample of the documents input to the particular week's run were printed out, duplicated and despatched for users to mark relevant items.

The first sets of random lists covered documents input to the matching run for week 031 and three sheets were despatched to each user. On the basis of the returns the Recall was calculated to be 39 percent for R1 documents and 21 percent for all documents marked (i.e. R1 and R2).

However, in view of the small number of sheets sent many users found none of the documents to be relevant. A second sample listing comprising six sheets was therefore sent to another set of users. This latter sample gave a Recall figure of 43 percent (R1 documents) and 24 percent (R1/R2 documents).

Further random listings were sent to users for the weeks 050 and 056. In the latter case both random listings and bulletin listings were sent for purposes of comparison. From this it would appear that the Recall figures derived from marking of random lists gives figures that are similar to those derived from bulletin markings assuming that the sample size is not too small.

The data derived from the various lists are set out in Figures 1 for bulletin listings and in Figure 2 for random lists. Data are also given for the subset of those profiles in the sample which remained unmodified throughout the investigation. In each case the figures given for the size of sample indicate the number of returns actually contributing to the figures.

From the data in the figures it appears that the overall Recall performance of the system was at a fairly high level at the beginning of the service i.e. for weeks 014 and 023, dropped fairly sharply for the weeks 031 and 050 and rose again to close to its earlier level at week 056.

FIGURE 1.

Recall performance of SDI profiles based on Bulletin listings

Run Number	Single Marking Recall	Sample Size	Recall R1	Sample Size	Recall R1/2	Sample Size
<u>014</u> Total unmodified subset	52 (53) (60)	200 125				
<u>023</u> Total unmodified subset			58 (56) 54 (55)	269 80	45 (43) 45	337 102
<u>056</u> Total unmodified subset			46 (49) (50)	72 17	35 (40) (40)	92 21

\* NOTE Figures without brackets are averages of ratios , figures in brackets are averages of numbers

FIGURE 2

Recall performance of SDI profiles based on random lists

Run Number	Single Marking	Sample Size	Recall R1	Sample Size	Recall R1/2	Sample Size
031 (3 Sheets)			39 (42)	33	21 (20)	71
031 (6 sheets)			42 (31)	47	24 (19)	101
Unmodified (3 sheets)	31 (25)	33	39	16	23	29
Modified (6 sheets)			47	14	30	28
050			30 (31)	40	24 (23)	60
056 (unmodified)			48 (49) (53)	60 14	32 (35) (27)	96 26

FIGURE 3.

Comparison of 023 and 056 recall using the same sample of users

1. Average of numbers

023 Recall	RT	R1/2
Users sent 056 bulletins	56	39
Users sent 056 random lists	55	42

056 Recall	RI	R1/2
Users sent 056 bulletins	48	40
Users sent 056 random lists	53	35

2. Average of percentages

023 Recall	RI	R1/2
Users sent 056 bulletins	61	45
Users sent 056 random lists	57	48

056 Recall	RT	R1/2
Users sent 056 bulletins	49	42
Users sent 056 random lists	53	41

Figure 4

BULLETIN SUBJECT HEADINGS AS USED FOR GENERAL CLASSIFICATION  
OF DOCUMENTS IN THE SDI INVESTIGATION

- 10 Geophysics, astrophysics, astronomy, radio astronomy
- 14 Plasmas, ionization and discharges in gases
- 16. Electric and magnetic fields, particle optics, ion sources, accelerators, electron guns
- 18 Vacuum technology
- 20 Solid state physics, crystal structure, electron states, acoustic, thermal, electrical properties and effects
- 30 Superconductivity, cryogenics
- 32 Magnetism, magnetic materials and properties
- 34 Dielectric, materials and properties, ferroelectricity and piezoelectricity
- 40 Optical properties of materials, luminescence, fluorescence
- 42 Quantum electronics, masers, lasers, holography
- 50 Semiconductor materials and devices, crystal growth, microelectronics
- 60 Electron tubes, cathode ray tubes, thermionic tubes, photomultipliers, conductors, inductors, resistors, capacitors, and switches
- 64 Circuit theory, network analysis and synthesis
- 66 Electronic circuits, amplifiers, modulators, oscillators logic circuits, pulse circuits, power supply circuits frequency dividers and multipliers
- 68 Reliability, quality control, testing
- 70 Telecommunications, radio, television, information and communication theory, signal processing
- 75 Radar
- 76 Antennas and propagation

- 77 Sonics and ultrasonics
- 78 Microwave technology
- 80 Electric machines, power conversion
- 81 Aerospace facilities and techniques, space communication
- 83 Direct energy conversion and energy storage
- 85 Instrumentation and measurement, biomedical engineering, telemetry
- 88 Particle and radiation measurement
- 90 Control theory and components, switching theory, artificial intelligence, cybernetics
- 96 Computer technology and applications

## Chapter 18

### ASSESSMENT OF THE RECALL PERFORMANCE OF THE SDI SYSTEM IN RELATION TO PARTICULARLY VALUABLE ARTICLES

#### Introduction

During the SDI Investiation it was desired to measure the Recall and Precision performance of the system. Of these Recall was, of course, the most difficult to assess.

The normal method was to present the user with a listing of a sample of the documents input to the system and to ask him to mark any item(s) which he considered should be selected by his profile. The marked items were then compared with those items actually notified to him and Recall was calculated as the ratio of items marked and notified to total documents marked.

There is, of course, nothing unusual in this method. The point is that figures derived in this way are likely to underestimate the Recall performance rather than to overstate it. The user was free to mark items of passing interest or ones with no direct relationship to the subjects included in his profile.

From these and other considerations it may be supposed that the actual Recall figures of the system based on those items which fall directly within the stated field of the profile could well be higher than the figures indicate. In particular the Recall performance for the important articles central to the user's subject field could well be considerably higher than for a wider sample which includes items of fringe interest.

A small study was undertaken to test this hypothesis.

## Method

A letter, Appendix 1, was sent to each participant in the SDI Investigation asking him for details of any papers published in the previous six months which he had found particularly valuable. To avoid any direct connection with the SDI Investigation the letter was signed by Mr. T M Aitchinson, Manager, Information Research, INSPEC: (it should perhaps be pointed out that since the beginning of the Investigation all direct communications with the users had been signed by the author as Manager, SDI Investigation). The working of the letter was intended to eliminate bias towards SDI notifications and as far as possible to prevent users being influenced in any way by the SDI service. These considerations naturally meant that we could not ask users to restrict themselves to English-language material or to the subject area of their SDI profiles.

The return form left space for three articles since only the really important items we required and in fact the possibility of even fewer items was suggested by asking for 'NIL' returns. However users who wished to include more than three articles were permitted to do so.

Some 450 individual participants were sent the questionnaire and 219 replies had been received by the time it was desired to examine the results. The accompanying letter made a reply a matter of choice so that the low percentage return (under 50 percent) is not surprising, particularly since the questionnaire was sent out at the height of the summer holiday period. However, it is not considered likely that these particular results would be seriously affected by a small percentage return.

As was to be expected only a proportion of the items cited had appeared among those input to the SDI system. The coverage of the Investigation extended only to English language articles in a number of chosen periodicals. Naturally some of the items cited had appeared in journals other than these, and yet other items were from strictly speaking, non-periodical sources, eg conference proceedings. In addition, though some attempt was made in the covering letter to restrict items to those published within the period of operation of the SDI service, some of the papers cited had been published before the start of the service in November 1968. For these and other reasons only 286 of the items cited had actually been input to the SDI Service and could therefore be used in this study. The number of individual returns contributing to this total was 126.

In using the remaining items it was realised that the validity of the results would depend on the extent to which the choice of items cited was biased towards those actually notified by the SDI Service.

The most serious case would be that in which the participant had intentionally restricted the scope of his selection to such articles. It was, however, possible to calculate the maximum extent of this bias and to make allowance for it by excluding any return comprised solely of items which had been included in the particular user's SDI notifications.

There is, however, one other type of bias which could affect the results. It may be supposed that for any person regularly receiving an SDI service many of the documents known to him will be those notified by the service. The subset of such documents selected by him as particularly valuable is therefore likely to be biased towards SDI-notified documents.

Little direct evidence could be derived from the sample to determine the effect of this second possible bias. It was clear from the date already listed before the corresponding SDI notifications were sent. These items formed a clearly unbiased sample, but one which was small, comprising only 17 articles:

In order to obtain a larger but similarly unbiased sample we took some of the cited items which had not been input to the system for various reasons. This sample included articles from journals not scanned for the service, articles from issues which had not been received originally as well as from issues predating the start of the service. Copies of these articles were obtained and passed through to the indexers in the usual manner as part of the regular input. The indexers were, therefore, unaware that these items were in any way special and there could, therefore, be no question of special treatment for these articles.

In all, some 37 articles were obtained and indexed in this way. A larger sample would have been preferred but the effort required to locate, obtain, photocopy and index the documents made this difficult.

With the 286 documents which had been input to the system normally these additional 37 articles gave a total sample of 323 documents.

In the case of the 686 documents it was possible to find whether they had been notified to the particular participant by inspecting the record of his notification. However, for the 37 items the only practical way was to do a manual matching of the index terms assigned to them against the particular profiles to discover if the documents would have been selected.

This method was somewhat laborious but avoided the confusion that might have been caused if the documents had been input to the system normally and had appeared in the system normally and had appeared in the weekly notifications.

In examining the profile to discover whether they would have selected the documents, it became clear that some of the documents bore little direct relation to the user's stated information requirements used as the basis for his SDI profile. This was not surprising since, in our efforts to obtain an unbiased sample of documents we had avoided any indication that the items cited should be restricted to any given field. This was not too serious since profiles in the Investigation had always attempted to cover the full interests of participants rather than a portion only.

However, it was apparent that some items were quite outside the scope of the profiles and some thought was given to removing them from the sample. This idea was abandoned owing to the difficulty of deciding unequivocally whether a particular article could properly be considered to be within the scope of an often fairly broadly defined Statement of Information Requirements, and it was decided to include all items and to accept that this would result in a somewhat lower Recall performance figure.

### Results

The results of the study are summarized in Table 1. As can be seen, of the sample of 286 documents 223 were selected by the SDI Service giving a Recall figure of 78 percent (average of ratios).

To assess the extent of bias caused by any participant intentionally restricting his choice only to items included in his SDI notifications the figures were recalculated to exclude any returns where all items had been included in the SDI Service and had been notified to that user. As these returns, by definition, produced figures of 100 percent Recall it was to be expected that their exclusion would considerably reduce the overall Recall figure.

However, it can be seen that the figures are very little lower, ie 74 percent (average of numbers) and 72 percent (average of ratios).

Using the 54 documents comprising the 37 specially-indexed documents and the 17 items included in the returns before they were notified to the users, it was possible to establish a Recall figure for the system on the basis of a sample which was clearly free of bias towards SDI-notified documents.

The figure of 61 percent (average of numbers) and 65 percent (average of ratios) can be taken as the "worst case" estimate for the Recall performance. As stated above a number of the documents in the sample lay outside the scope of the profiles and could reasonably have been excluded from the sample with the result of increasing the calculated Recall for the remaining documents. The other factor tending to a lower Recall performance is that, at the time these documents were indexed, a study was being carried out of the effect on system performance of reducing the number of terms allocated in document indexing. As can be seen from Table 1 the reduced indexing resulted in a considerable fall in Recall performance and must, undoubtedly, have reduced the Recall for this particular sample. The end two columns of Table 1 show the nearest comparable Recall figures for documents in general calculated for both highly relevant (R1) documents alone and for highly relevant and partially relevant documents combined (R1/2). It will be seen that in the period of both normal and reduced indexing the Recall performance for the particular valuable documents is considerably higher than for documents in general.

### Conclusions

It appears clear from the results that the Recall performance of the SDI system is considerably higher for those items which recipients would consider most valuable, than for items in general. The figure of 75 percent obtained in this study could well be higher if items outside the scope of the given profiles were excluded from the sample.

It appears likely that a reduction in Recall performance affects the highly relevant documents proportionately less than the less relevant ones.

TABLE 1

Sample	Number of Documents	Number Notified	Particularly valuable documents		All documents	
			<u>RECALL</u>		<u>RECALL</u>	
			Average of Numbers	Average of ratios	R1	R1/2
Original 286	286	223	78	76	58	45
Reduced 125	125	92	74	72	58	45
'Unbiased' sample	54	33	61	65	43	24

## Chapter 19

### SUBJECTIVE ASSESSMENT OF SDI SYSTEM

#### RECALL PERFORMANCE

##### Introduction

Throughout the SDI Investigation major effort was devoted to measuring the performance of the system. However it was at all times realised that in any operational service the most immediately important measure of performance is the user's degree of satisfaction with the service.

##### Method

In the final questionnaire sent to participants at the end of the Investigation an attempt was made to discover the user's assessment of the service they had received and the extent to which they considered it had been successful in picking up the relevant items in their field.

Two questions gave us information on the user's assessment of the Recall performance of the service. These were

1. How many English-language periodical articles published over the last year did you find particularly valuable in your work?
2. Of these how many were notified to you by the SDI service?

By no means all the respondents to the questionnaire answered these questions, a small but significant proportion obviously shying away from answering questions for which they had no factual answers. However usable replies were received from almost 300 people.

## Results

The number of documents given by individual participants varied widely ranging from 0 to 300. The results are tabulated in Figure 1 and show the Recall figures for users in each of the user groups-Universities, Government establishments, Industrial firms and Group profiles.

It is noticeable that in spite of the large variation in numbers of documents given in individual returns, the percentage Recall figure calculated as average of numbers agrees very clearly with that calculated as average of ratios.

It will be seen that the Recall figure for University users is higher than for Industrial users which again is higher than for Government users. Determination of the significance of these differences requires further work but the difference between University and Government users appears on the face of it too large to be due to chance. The low figure for Group users should be treated with caution in view of the small number of people in this group.

Figure 1

### Subjective assessment of SDI System Recall Performance

Type of user	No of returns	Total particularly valuable articles	Total notified by SDI	Recall %	
				Av. of Nos.	Av. of ratios
UNIVERSITY	76	2023	1367	68	68
GOVERNMENT	93	2163	1264	58	57
INDUSTRY	91	3053	2024	63	65
GROUPS	22	1131	602	53	53
TOTAL	282	8370	5257	63	62

## Chapter 20

### RELATION BETWEEN PRECISION PERFORMANCE OF THE SYSTEM AND DOCUMENT INDEXING

For any information retrieval system the performance must be dependent to some extent on the indexing. The most easily isolated feature of the indexing is the average number of terms assigned to each document that is added to the system.

As the number of index terms per document decreases, it may be assumed that the smaller number of terms will tend to cover the major rather than the minor concepts in the documents: it follows that documents retrieved by these terms have a greater probability of being relevant. From a different standpoint, if the number of document index terms is reduced, the number of possible retrieval 'hooks' is reduced, the number of documents pulled out of the file is less, giving a probability of reduced Recall and therefore of increased Precision.

It was expected therefore that an inverse relationship would be found between the average number of index terms assigned per document and the Precision performance of the system. A graph was plotted of this relationship and the result is shown in Figure 1. The result is not as clear as one would wish, the scatter of precision values for the same number of index terms being very large.

In view of this it was decided to look at one of the assumptions underlying the supposed relationship between index terms and Precision.

It is assumed that a) a reduction in the number of index terms will cause a reduction in the number of documents selected by the user's profile, and b) the smaller number of documents will contain a higher proportion of relevant ones. Only if a) is true can b) be true.

A plot was therefore made of the relationship between average number of index terms and the average number of notifications sent to each user. The result is shown in Figure 2: the points relate to 4-week periods, the number against each point indicating the first week of each period. Again, however, it can be seen that a clear relationship cannot be derived and it would appear that a third factor is involved. Examination of the data in Figure 3 indicates that this may be the number of documents input to the system in each of the four-week periods.

Figure 4 shows a plot of total documents input to each four-week period against average number of notifications. It can be seen that the points for periods 11, 15 and 19 are isolated and these relate to periods where the number of documents input was relatively low.

Finally the data in Figure 1 was again plotted for the four-weekly periods. As can be seen from Figure 5 the result is a fairly satisfying curve when the precision figures used are calculated as average of numbers but in Figure 6 it can be seen that averages of percentages give one or two anomalous points.

### Conclusions

It may be concluded from the results that, in the system used for the SDI investigation:

- 1) As expected, the number of notifications produced for a user in any run is directly related to the number of documents input to that run.
- 2) The number of notifications is directly related to the average number of index terms applied to each document in the run. The equation for the line drawn in Figure 2 which has been somewhat arbitrarily chosen as passing through the origin is  $y = \frac{6x}{5}$

- 3) The Precision performance is indirectly related to the average number of index terms applied to each document in the particular run. The equation for the line drawn in the graph (Figure 5) is  $y = \frac{4}{3}x + 79$ .

This indicates that the Precision performance of the system is surprisingly little affected by the average number of index terms assigned to each document and that a figure of approximately 80% Precision is the highest attainable by means of more restricted indexing.

At the same time the number of notifications per user is much more affected by the average number of index terms. Since a considerable drop in number of notifications is not accompanied by any commensurate increase in the percentage of relevant notifications it must follow that the percentage Recall performance suffers seriously. At the same time it would also appear that in this system high Recall was more readily obtainable with good Precision than High Precision with good Recall.

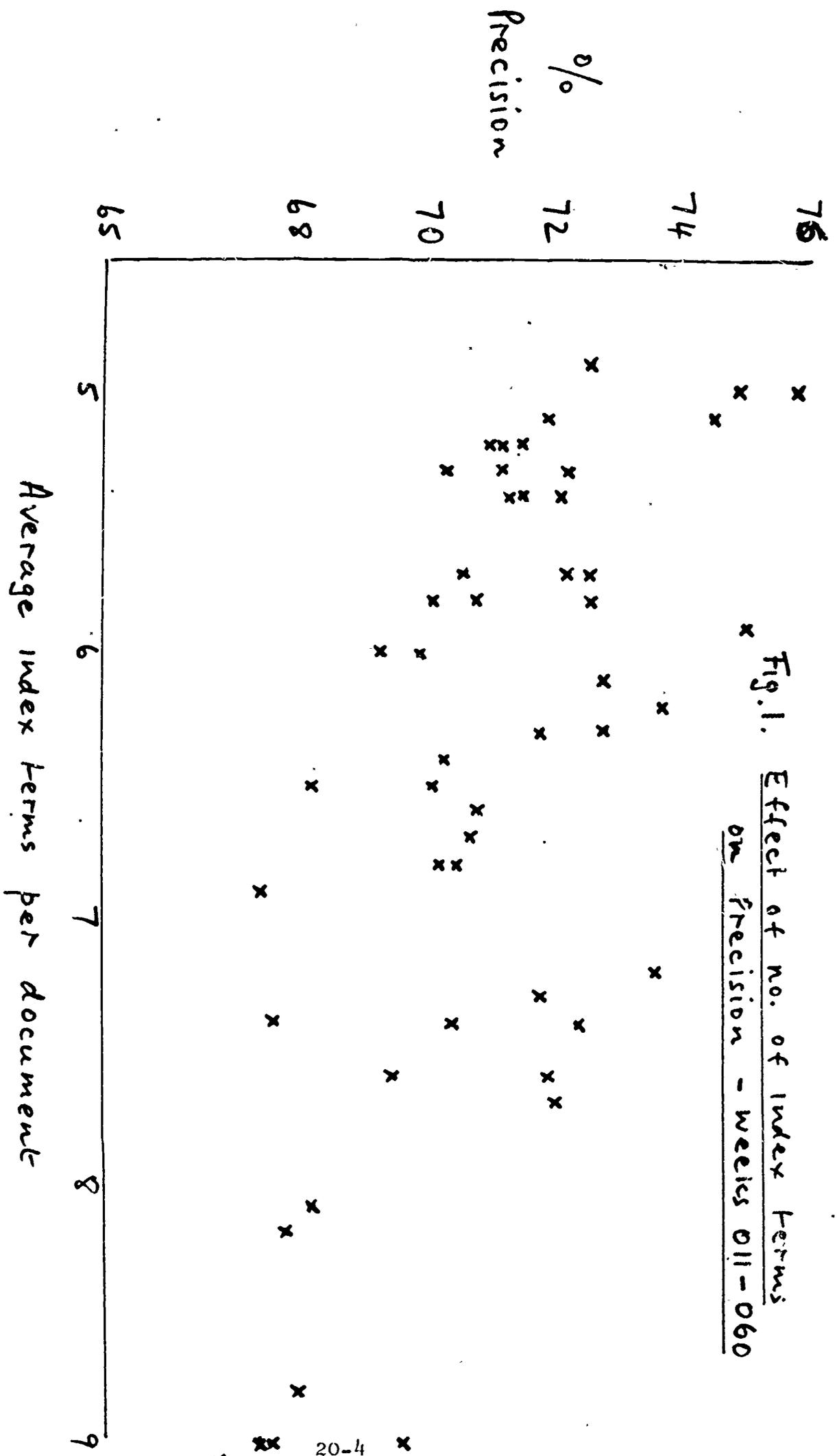


Fig.2. Relationship between no. of notifications and average index terms per document.

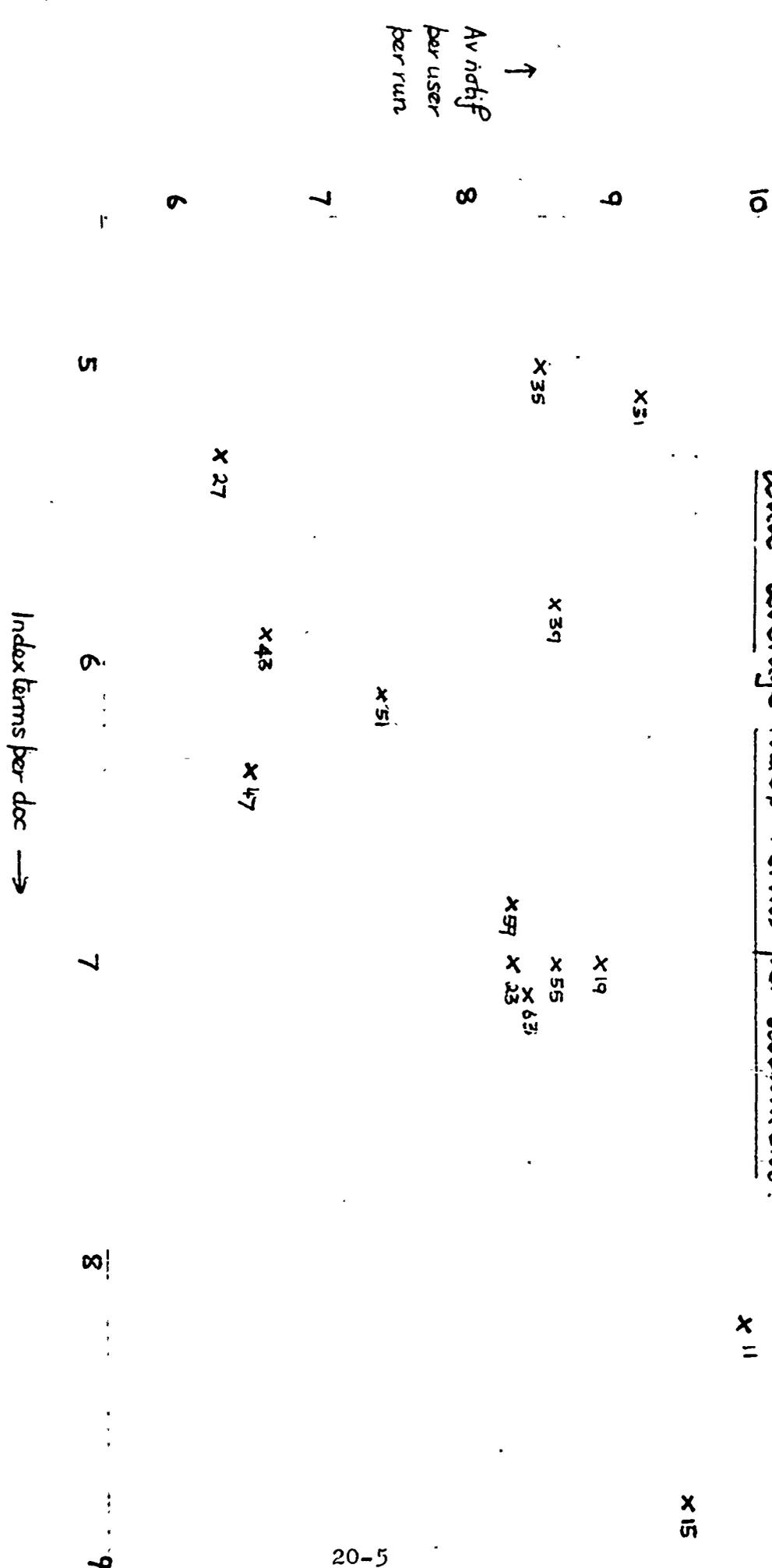


FIGURE 3.

Performance figures relating to 4 week periods

Run Nos.	Total Docs,	Index Terms per doc.	Total no. of index terms	No. rel. No. sent	Average Precision	Av. notif. per user per run.
011-014	923	8.2	7566	69.7	69.5	9.9
015-081	989	8.8	8653	67.7	66.5	9.5
019-022	858	7.0	6021	71.4	69.0	8.9
230-026	1135	7.0	7910	69.8	69.3	8.3
027-030	988	5.3	5269	71.3	70.5	6.3
031-034	1493	5.1	7138	72.6	71.4	9.2
035-038	1311	5.0	6618	73.3	71.6	8.5
039-042	1299	5.8	7485	71.4	70.5	8.6
043-046	962	5.9	5657	72.7	70.4	6.6
047-050	980	6.3	6151	71.4	69.1	6.5
051-054	928	6.1	5667	72.1	69.2	7.4
055-058	1141	7.0	7988	69.2	65.5	8.6
059-062	1108	6.8	7576	71.9	69.0	8.3
063-066	1167	7.1	8242	70.0	67.3	8.4

Fig. 4. Average number of notifications per user  
vs. total documents input.

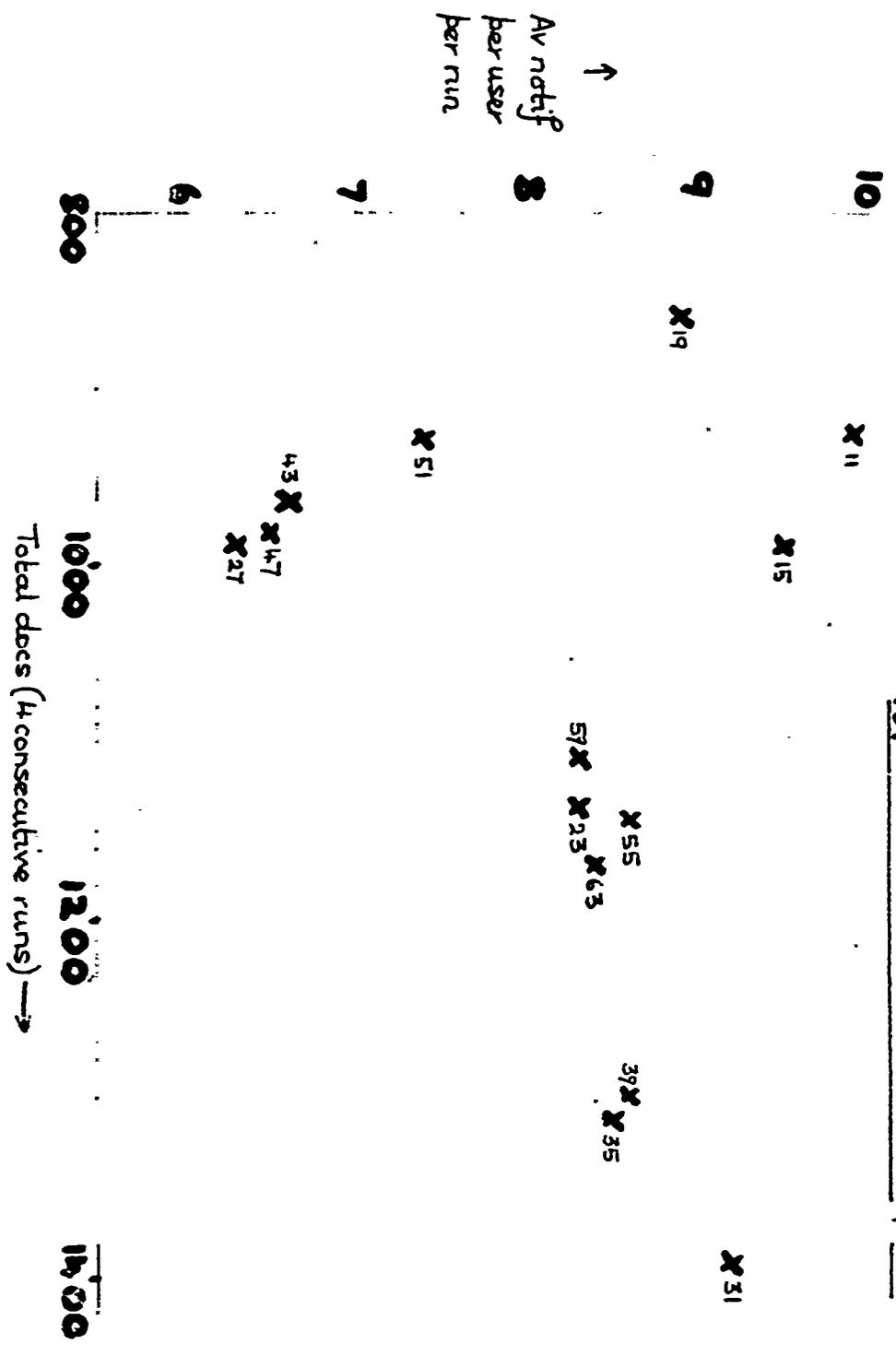


Fig. 5. Precision performance vs average no. of index terms per document.

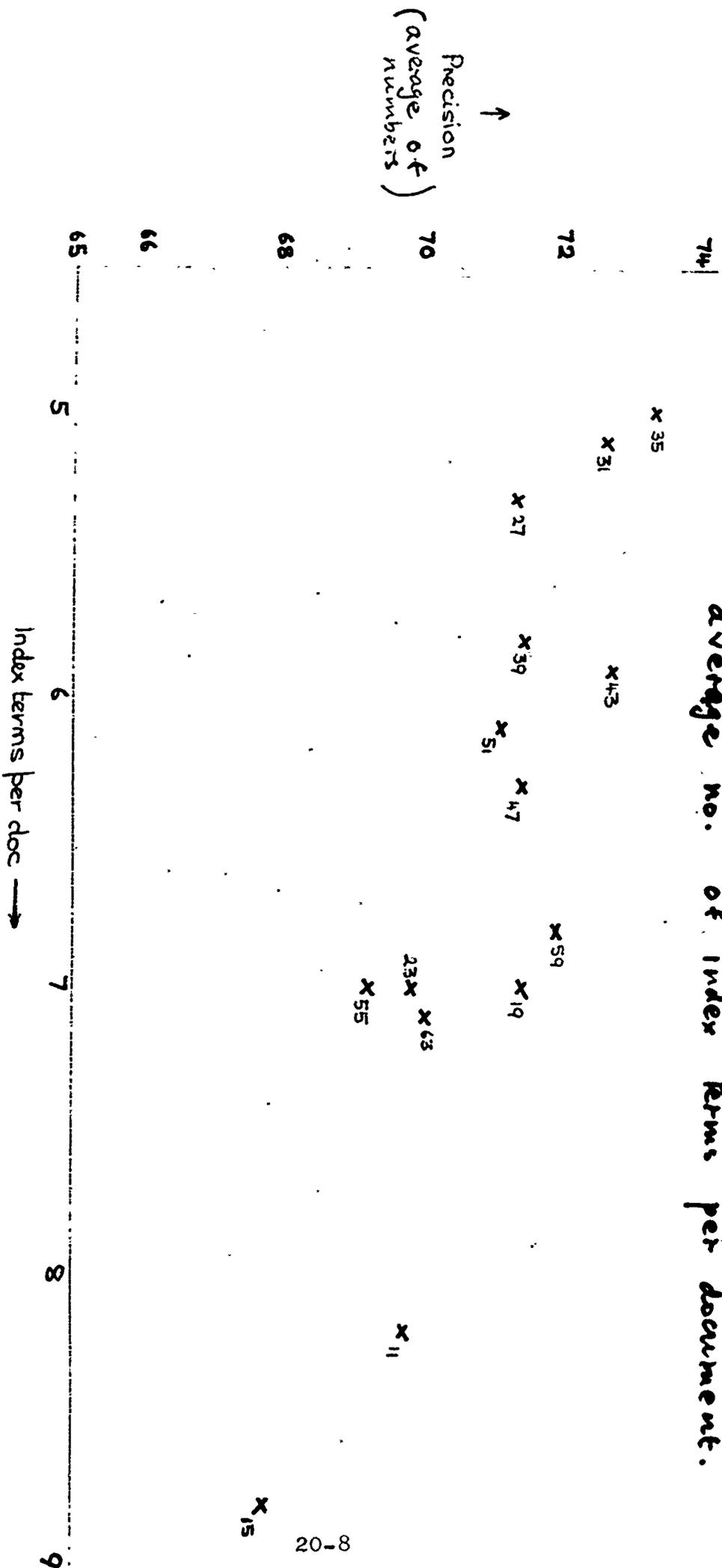


Fig. 6. Precision performance (average of ratios) vs average no. of index terms per document.

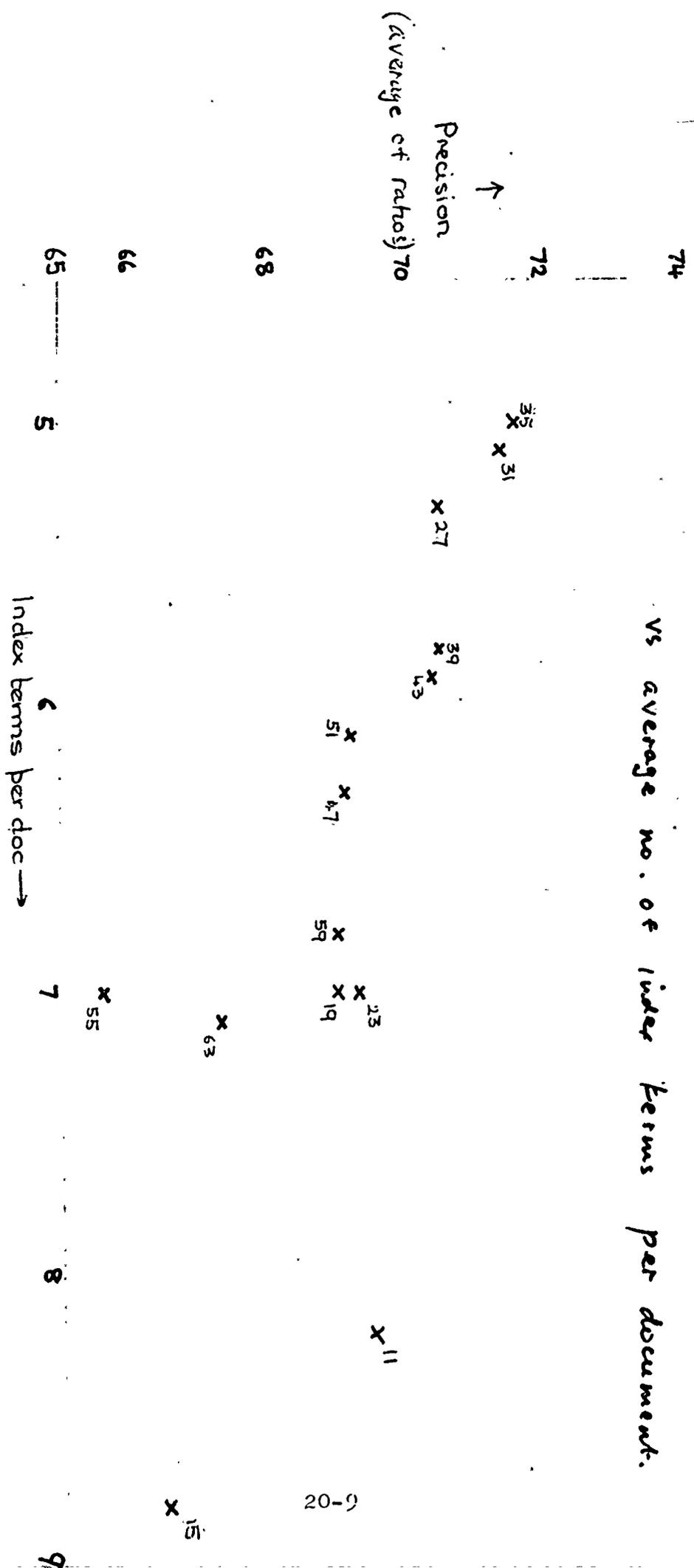
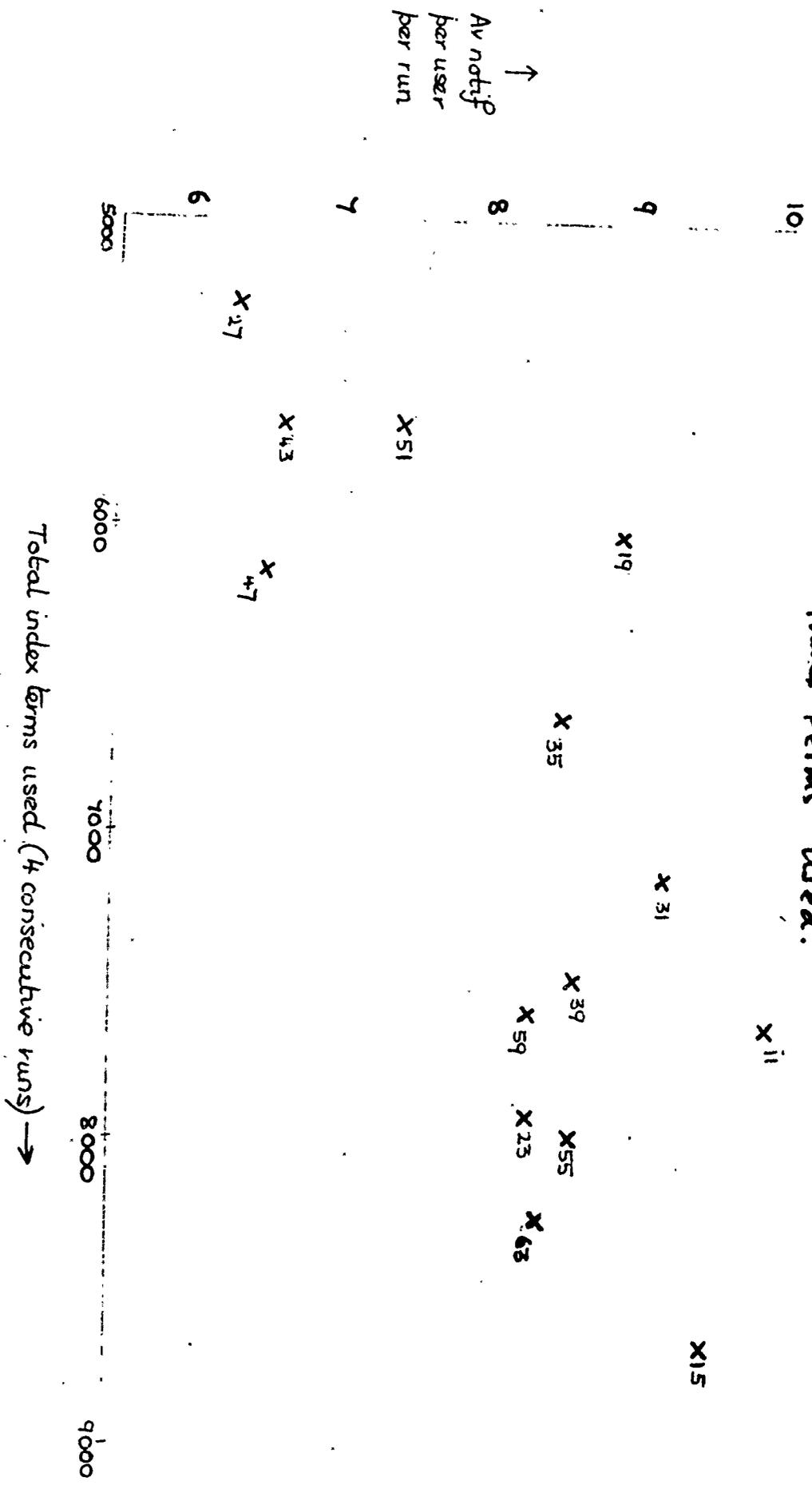


Fig. 7 Average notifications received per user vs total index terms used.



## Chapter 21

### USER ASSESSMENT OF THE SDI SERVICE

One of the main problems in the SDI Investigation was that of monitoring the performance of more than 500 profiles, diagnosing problems and acting quickly to prevent any user becoming disenchanted with participation in the Investigation. The problem common to all SDI systems is that the overall performance of the service has no meaning for the individual user who sees only the performance of his own profile. The chief virtue of the SDI service in meeting the specific individual needs of subscribers becomes its chief disadvantage when it performs badly, since the failure is felt directly, personally and repeatedly week after week unless something is done to improve matters.

The weekly service to participants in the Investigation started in November 1968 and it was realised that it was desirable to discover quickly which recipients were finding the service unsatisfactory so that the limited staff effort could be devoted to the most serious cases. After eight weeks' service therefore, each participant was sent a questionnaire asking his opinion of the service and which aspects he found least satisfactory. The assessment categories were Very Poor, Poor, Fair, Good and Very Good. The returns are summarized in Table 1.

In the final questionnaire at the end of the Investigation when participants had been receiving service for approximately fourteen months the same question was included. The results of this second questionnaire are shown in Table 2.

It can be seen that the overall assessment of the service had improved with approximately 72 percent rating the service Good or Very Good compared with 51 percent in the previous questionnaire. Less than five percent considered the service Poor or Very Poor compared with twelve percent earlier.

However, not all users who replied to the first questionnaire replied to the second and vice versa. Tables 3 and 4 give the results for the two questionnaires for the people who replied to both. These results therefore allow a direct comparison.

It can be seen from these figures that the proportion of people rating the service Good or Very Good had risen to 72 percent compared with 54 percent in the earlier questionnaire and only five percent thought it Poor or Very Poor compared with 11 percent in the earlier questionnaire. These figures agree very well with those obtained from Tables 1 and 2 and indicate that the sample of users replying to the second questionnaire was not significantly different in kind from that replying to the first.

It may be concluded that during the course of the Investigation the overall degree of user satisfaction with the service improved. However, it is of some interest that this increase in satisfaction was not common to all users. Table 5 shows the number of people whose stated degree of satisfaction differed between the first and second questionnaire. It can be seen that although 38 percent (110) of the users rated the service better in the second questionnaire, the degree of satisfaction of 10 percent (29) of the users had decreased.

#### Assessment compared with subjective recall performance

In the final questionnaire, as well as giving an overall assessment of the SDI service, users were asked to state how many of the relevant articles in their field had been notified to them by the SDI service. It was, therefore, possible to compare this subjective Recall performance assessment with the degree of satisfaction with the service.

The results are shown in Table 6 for 272 users. It can be seen that the Subjective Recall performance figure (72 percent) for those people who considered the service Good or Very Good was higher on average than that (45 percent) for those who considered it Fair. This, in turn, was higher than the figure of 34 percent for those who rated the service as Poor.

This is, of course, what one would expect but it does serve to show an agreement between these two subjective judgments.

Table 1 User assessment of SDI Service after two months

Type of User	TOTAL USERS	VP	P	F	G	VG
University	133	3 (2.3%)	12 (9.0%)	49 (37%)	54 (41%)	15 (11%)
Government	148	1 (0.7%)	18 (12%)	55 (37%)	66 (45%)	8 (5%)
Industry	122	5 (4.1%)	11 (9.0%)	44 (47%)	57 (47%)	5 (4.1%)
ALL	403	9 (2.0%)	41 (10%)	148 (37%)	177 (44%)	28 (7%)

Table 2 User assessment of SDI Service after fourteen months

Type of User	TOTAL USERS	VP	P	F	G	VG
University	98	1 (1.0%)	4 (4.0%)	20 (20%)	50 (51%)	23 (23%)
Government	113	-	8 (7.1%)	29 (26%)	60 (53%)	16 (14%)
Industry	113	1 (0.9%)	1 (0.1%)	28 (25%)	65 (58%)	18 (16%)
ALL	324	2 (0.6%)	13 (4.0%)	77 (24%)	175 (54%)	57 (18%)

Table 3 Assessment of service by sample of 287 users after two months

Type of User	TOTAL USERS	VP	P	F	G	VG
University	88	1	5	33	42	7
Government	103	1	13	36	45	8
Industry	96	3	8	32	50	3
ALL	287	5 (1.7%)	26 (9.1%)	101 (35%)	137 (48%)	18 (6.3%)

Table 4 Assessment of Service by sample of 287 users after 14 months

Type of User	TOTAL USERS	VP	P	F	G	VG
University	88	1	4	17	46	20
Government	103	-	7	28	53	15
Industry	96	1	1	20	60	14
ALL	287	2 (0.7%)	12 (4.2%)	65 (23%)	159 (55%)	49 (17%)

Table 5 SDI recipients giving a different assessment of the Service in the second questionnaire compared with the first questionnaire (Sample of 287 users)

	Higher Grading on Second Questionnaire			Lower Grading on Second Questionnaire		
	By 1 Grade	By 2 Grades	By 3 Grades	By 1 Grade	By 2 Grades	By 3 Grades
University	30	6	-	8	1	1
Government	28	4	1	11	-	-
Industry	31	10	-	8	-	-
	89	20	1	27	1	1

Table 6 User rating of SDI Service compared with their  
assessment of Recall performance

		Very Poor	Poor	Fair	Good	Very Good
University	No. of Users Total % Recall Average %	-	3 150 50	17 845 50	41 2964 72	15 1178 79
"	No. of Users Total % Recall Average %	-	8 159 20	25 1093 44	49 3331 58	13 788 61*
Government	No. of Users Total % Recall Average %	1	1 100 100	22 967 44	56 4059 72	11 784 71
Industry	No. of Users Total % Recall Average %	1	12 409 34	64 2905 45	146 10345 71	39 2750 71
ALL		-				

\* These include a return from one user who rated the service Very Good in spite of the fact that, by his figures, the Recall performance was O. This is explained by the fact that he included only two items as being of particular value, both being outside the scope of the service. One was an unpublished communication and the other a paper given at a conference in Germany.

## Chapter 22

### RELATIONSHIP BETWEEN QUALITY OF USER'S INITIAL STATEMENT OF INFORMATION REQUIREMENTS AND THE SUBSEQUENT PERFORMANCE OF HIS PROFILE

With increasing experience of profile compilation and performance testing, a feeling grew that certain types of user statements were more likely to produce satisfactory profiles than others. Such a feeling might merely reflect a greater degree of confidence on the part of the SDI staff dealing with certain types of requirements and it might not follow that the resulting profile gave greater satisfaction to the particular user.

A study was therefore made to discover whether particular statements considered likely to give good performance actually did in practice perform better than others.

#### Method

A sample of 80 users was selected and an examination made of the original statements used for profile compilation. Based on this examination an assessment was made of the quality of the statement as Good, Poor or Fair i.e. a prediction of the performance of the resulting profile.

The factors user in assessing the statement were:-

POOR CHARACTERISTICS

1. Statement

a) Lack of detail

Examples of statements lacking detail are 'I am interested in nuclear instrumentation' or 'Please inform me of articles on airborne radar equipment'. Each is sufficient to define the general area of interest but gives insufficient information for compiling a profile to select articles in the particular part of the field that is likely to be of interest to the user.

b) Lack of precision

Lack of precision generally arises from failure of the user to realise that a greater logical precision than customary is needed in stating his requirements for the purposes of a search profile. Thus it is common to receive statements of the form 'I am generally interested in semi-conductors of the III-V group but particularly GaP, InP and InSb'. Without specifying those particular aspects of the III-V group of compounds in general which are of interest, the profile can only opt to cover either all of the III-V compounds or just the three specifically-mentioned.

The problem is not always easily resolved even by reference to the user since he has often not consciously formulated the principles on which he would select material himself.

2. Subject Field

a) Difficulty of coding.

Some subjects have a much more settled and agreed vocabulary than others and coding of search profiles is therefore easier for these subject areas than for others where the vocabulary is less defined. Examples of areas where difficulty in coding was found included control, information theory and man-machine systems.

b) Fringe interests

A problem with any SDI system is how to decide which user interests can be covered by the subject coverage of the data base. In many cases an individual's interests will fall across a number of subject areas, some of which will be fully covered by the particular data base, some partly and some not at all. For such fringe areas the performance (particularly Recall) of the profile and therefore the users assessment are likely to be poor.

GOOD CHARACTERISTICS

These are the reverse of the characteristics enumerated above. Examples of statements with good characteristics in each category can be seen from Table 1.

1. Statement
  - a) Plenty of detail
  - b) High precision
2. Subject Field
  - a) Ease of coding

FAIR CHARACTERISTICS

Statements which could not be categorized according to any of the above were put in the Fair category.

In practice it was found that many statements were assigned to 2 or 3 categories. In those cases where there was a mixture of good and poor characteristics, prediction of performance was based on a simple balancing out of the categories.

The actual performance against which the prediction was compared was that based on the Test Collection of documents. Performance figures were also available for later runs (viz. the experimental service runs 001-004, 005-006, and 014) but many of these were of profiles which had undergone modifications as a result of profile analysis or user interaction and this could be significantly different from their original state as compiled from the statements. These experimental service performance figures were in fact used for comparison with the user ratings of profile performance (obtained by questionnaires sent out around run 020); this comparison served as a measure of whether performance figures corresponded to user ratings.

In order to have a single overall figure for profile performance the product Recall X Precision was used, ie a profile operating at 60 percent Recall and 80 percent Precision would have a performance figure of 0.48 (0.6 x 0.8). A major disadvantage of this method is that it assumes the equivalence of Recall and Precision, eg profiles operating at one hundred percent Recall and 20 percent Precision, or 20 percent Recall and one hundred percent Precision both have the same performance figure but are obviously not equally satisfactory. However for the purposes of ranking profiles this method was considered convenient.

## RESULTS AND DISCUSSION

The complete details of the eighty profiles are shown in Table 1. Table 2 includes the main results: with the profiles listed in order of the Test Collection performance figure.

Copies of the statements for the sample and other profiles are given in Appendix 22A.

The relationships of the following were investigated:

- (1) Prediction of performance and Test Collection performance
- (2) User rating and the 001-014 experimental service performance
- (3) Prediction of performance and the 001-014 experimental service performance

In each case the average performance figure was calculated for the profiles in each performance category (see Table 3) and plotted against the performance category (see Figure 1 curves 1, 2 and 3).

In case 2, the User ratings, VP and P , and , G and VG, were combined in order to make curve 2 more directly comparable with the other two.

It should be noted that profiles with performance figures of 0 were ignored for purposes of computing averages as they are quite anomalous and usually arise from too little data.

## CONCLUSIONS

- (1) In general the performance predicted on the basis of the statement of information requirements was reasonably reflected by the results obtained on the test collection performance (curve 1) but much less so by the results obtained in the experimental service (curve 3) when further user/system interaction had taken place
- (2) The user's rating of the service reflected the performance of the service he received

### Longer-term effects

However it is possible that this higher performance and satisfaction is a short-term effect based on highly specific short-term requirements and that over the longer-term the more general statement may with the effect of interaction and feedback perform as well or better.

To study this an examination was made of the overall performance figures for the 12 months service period and of the user's satisfaction with the service as stated in the second questionnaire at the end of this time.

Table 4 shows for each of the profiles the user's assessment of the service as stated in the second service questionnaire and the overall R1 and R1/2 Precision performance for the 12 months service.

In figure 2 are shown the average Precision figures for each of the groups of profiles derived from statements graded as Poor, Fair and Good respectively. This would appear to show that at least in the case of the Good statements the effect of an initial statement on subsequent performance is lasting and not removed by subsequent interaction with the user and by profile modification.

However the sample is not of such a nature to allow firm conclusions to be drawn.

If one looks at the relationship between the predicted performance and the user's assessment of performance at the time of the first and second service questionnaire a number of points emerge.

Firstly it can be seen from figure that the prediction is not very accurate or at least does not coincide clearly with the user assessment since in only 18 (37%) profiles is there agreement, counting Good and Very Good together. However in most cases the prediction errs on the conservative side, since 39 (80%) of profiles are assessed the same as the prediction or higher and only 9 (20%) are assessed lower than the prediction.

Taking the profiles in the groups predicted 'Poor', 'Fair' and 'Good' respectively, of the 11 profiles in the 'Poor' group, 3 (27%) are assessed as Poor, while 8 (73%) are assessed as Fair or Good. Of the 17 'Fair' predictions only 3 (18%) are assessed as Fair while 13 (76%) are assessed as Good or Very Good. In the case of the 20 'Good' predictions 12 (60%) are assessed as Good or Very Good while 8 (40%) are assessed as only Fair.

It would appear from this that the predictions are not a very good guide to performance as interpreted by the user, though there appears a greater probability that a profile predicted to perform badly will in fact do so than those which are predicted to perform well.

However, since 64 percent of the profiles predicted to be Poor in fact get a Good or Very Good rating compared with 76 percent for those assessed as fair and 60 percent for those assessed as Good, it may be that factors likely to contribute to poor performance are more easily distinguished than those which contribute to good performance.

For the second questionnaire assessments the rating of the 'Poor' predicted profiles is again Good or Very Good in 7 cases (64%), while the figure for 'Fair' predicted profiles has risen to 14 (82%) and for the 'Good' to 18 (90%).

An examination of the 9 profiles which are still rated Poor or Fair at the time of the second questionnaire shows no clear pattern in the statements from which they were compiled, except that they covered subjects which lay somewhat on the fringe of the subject coverage of the service. These were 'nuclear magnetic resonance', 'man-machine systems' and environmental testing.

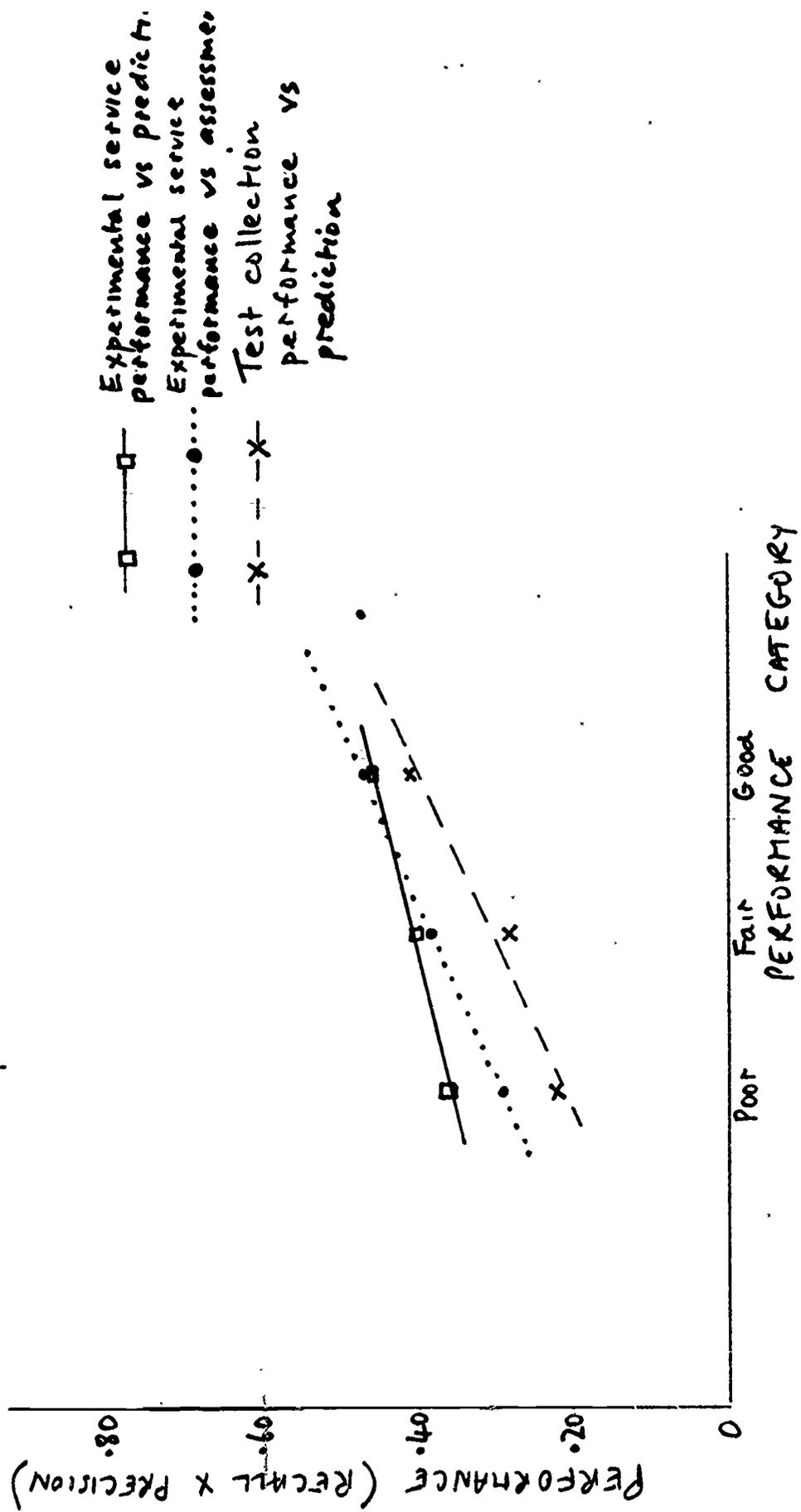
### CONCLUSIONS

The evidence for believing that recognisable characteristics of statements of information requests will predictably affect the performance of a profile is not as clear as one might suppose. The reason for this may be that the differences between statements obtained in the investigation were relatively small owing to the fact that users tended to follow closely the sample statement provided to them.

From the point of view of the performance figures themselves, profiles predicted to perform well or badly do tend to show different performances but this difference tends to become less with time, probably because interaction with the user compensates for initial defects.

The relationship between the performance predicted for a profile and the user assessment of its performance is not clear. In general the group of profiles predicted to perform badly are assessed lower than the group of profiles predicted to perform well. However a large number of profiles predicted to perform badly are assessed as Good by the user and vice versa. Profiles covering subjects on the fringe of the subject coverage of the service feature among those which perform badly as expected. At the same time not all profiles which are on the fringe of the field do in fact perform badly. The reason may be the manner in which the users interests in the field are made up - in some cases being more closely identified with the aspects dealt with in journals covered by the service.

Fig. 1. Average performance figures for profiles assessed or predicted to be Poor, Fair or Good



PROFILE No.	USER STATEMENT CATEGORIES						PREDICTION of PERFORMANCE	ACTUAL PERFORMANCE										USER RATING			
	GOOD			POOR				TEST COLLECTION			EST 001-006			OVERALL PERFORMANCE							
	1a	1b	2a	1a	1b	2a		2b	RECALL %	PRECISION %	PERFORMANCE FIGURE	R	P	F	R	P	R		P	RECALL %	PRECISION %
001	•		•				G	$\frac{1}{21} = 5$	$\frac{2}{4} = 50$	.03								$\frac{45}{70} = 64$	$\frac{64}{78} = 94$	.60	G
006				•			P	$\frac{0}{2} = 0$	$\frac{0}{2} = 0$	0								$\frac{14}{21} = 67$	$\frac{25}{38} = 66$	.44	F
013	•			•			F	$\frac{17}{35} = 49$	$\frac{29}{28} = 71$	.35								$\frac{41}{51} = 80$	$\frac{81}{98} = 83$	.66	VG
022							G	$\frac{1}{1} = 100$	$\frac{1}{3} = 33$	.33								$\frac{7}{9} = 78$	$\frac{11}{13} = 85$	.66	F
031	•						G	$\frac{9}{10} = 90$	$\frac{14}{14} = 100$	.90								$\frac{27}{55} = 48$	$\frac{32}{44} = 73$	.35	G
036				•			G	$\frac{6}{12} = 50$	$\frac{6}{10} = 60$	.30								$\frac{19}{16} = 63$	$\frac{15}{21} = 71$	.45	G
043							F	$\frac{2}{4} = 50$	$\frac{2}{4} = 50$	.25								$\frac{23}{60} = 38$	$\frac{48}{78} = 62$	.24	G
052				•			P	$\frac{19}{32} = 59$	$\frac{22}{40} = 55$	.32								$\frac{43}{93} = 46$	$\frac{58}{77} = 75$	.35	-
056	•						G	$\frac{0}{2} = 0$	%	0								$\frac{7}{7} = 29$	$\frac{2}{2} = 100$	.29	P
065	•			•			P	$\frac{4}{5} = 27$	$\frac{8}{11} = 73$	.20								$\frac{28}{108} = 26$	$\frac{49}{64} = 63$	.16	G
073				•			G	$\frac{4}{6} = 67$	$\frac{6}{12} = 50$	.34								$\frac{25}{35} = 71$	$\frac{34}{44} = 77$	.55	VG
082							F	$\frac{11}{14} = 79$	$\frac{22}{40} = 55$	.43								$\frac{17}{26} = 65$	$\frac{38}{114} = 33$	.21	F
089	•			•			F											$\frac{8}{22} = 36$	$\frac{18}{30} = 60$	.22	P
097							P	$\frac{3}{16} = 31$	$\frac{6}{7} = 86$	.27								$\frac{36}{74} = 49$	$\frac{52}{79} = 66$	.32	G
102							G											$\frac{3}{7} = 43$	$\frac{9}{15} = 40$	.17	F
107	•						G	$\frac{3}{5} = 60$	$\frac{3}{3} = 100$	.60								$\frac{0}{3} = 0$	$\frac{1}{6} = 17$	0	-
113				•			P	$\frac{1}{1} = 100$	$\frac{3}{30} = 20$	.20								$\frac{2}{8} = 25$	$\frac{1}{4} = 100$	.25	P
119				•			P	$\frac{4}{6} = 9$	$\frac{6}{7} = 86$	.08								$\frac{28}{28} = 100$	$\frac{62}{115} = 54$	.54	-
126	•						G	$\frac{9}{10} = 56$	$\frac{11}{11} = 100$	.56								$\frac{41}{95} = 43$	$\frac{54}{54} = 95$	.41	F
132				•			G	$\frac{9}{17} = 35$	$\frac{6}{6} = 100$	.35								$\frac{31}{38} = 82$	$\frac{27}{47} = 79$	.65	G
137				•			G	$\frac{5}{7} = 72$	$\frac{5}{6} = 84$	.60								$\frac{4}{7} = 57$	$\frac{8}{10} = 80$	.46	-

TABLE 1 (continued)

PROFILE No.	USER STATEMENT CATEGORIES				PREDICTION of PERFORMANCE	TEST COLLECTION			ACTUAL PERFORMANCE												USER RATING		
	GOOD		POOR			RECALL %	PRECISION %	PERFORMANCE FIGURE	EST 001-004			EST 001-006			EST 001-006 and Service Run 014			OVERALL PERFORMANCE FIGURE					
	1a	1b	2a	2b					R	P	R	P	R	P	R	P	RECALL %		PRECISION %	PERFORMANCE FIGURE			
	1a	1b	2a	2b																			
145					P					1/5	7/8			10/11	2/4	5/5	9/17	53	22/24 = 92	.49			
152															1/3	1/1	1/3 = 33			1/1 = 100	.33		
159					F	0/1 = 0	4/6 = 67	0		5/7	1/13	1/3	7/5	7/2	5/6	8/12 = 67	15/24 = 75			15/24 = 75	.50		
165					F	8/18 = 44	9/12 = 75	.33		23/28	29/32				6/7	9/9	29/35 = 83	35/41 = 85			35/41 = 85	.71	G
170					F	3/4 = 75	9/17 = 35	.26		1/5	5/11	1/4	3/9	4/5	3/6	6/4 = 43	1/2 = 42			1/2 = 42	.18	F	
175					F	4/20 = 20	4/10 = 40	.08		15/24	32/47			2/6	8/11	17/30 = 57	40/58 = 69			40/58 = 69	.39	G	
194					F	9/14 = 64	14/20 = 70	.45		5/10	16/23	8/11	13/26	4/5	7/8	17/26 = 65	36/57 = 63			36/57 = 63	.41	G	
199					G	9/18 = 50	9/23 = 39	.20		30/83	42/51			8/13	9/14	38/66 = 58	51/78 = 78			51/78 = 78	.45	VG	
204					G	5/9 = 56	12/14 = 86	.48		8/17	17/23	3/8	4/10	0/1	4/11	11/26 = 42	25/44 = 57			25/44 = 57	.24	VG	
209					F	2/7 = 29	3/4 = 75	.22		8/9	15/22					8/9 = 89	15/22 = 68			15/22 = 68	.61	G	
215					F	0/2 = 0	1/1 = 100	0		2/4	4/12					2/4 = 50	1/2 = 33			1/2 = 33	.17	F	
220					G	3/7 = 43	3/3 = 100	.43		8/19	8/9	10/13	14/19	2/3	2/2	20/35 = 57	24/30 = 80			24/30 = 80	.46	G	
230					P	0/2 = 0	0/1 = 0	0		5/7	9/6	3/6	7/7	2/4	5/5	10/17 = 59	18/18 = 100			18/18 = 100	.59	G	
236					G					5/10	5/5			1/3	2/3	6/13 = 46	7/8 = 88			7/8 = 88	.40	G	
242														2/2	2/2	2/2 = 100	2/2 = 100			2/2 = 100	1.0	G	
246					F	12/37 = 32	13/20 = 65	.21		33/59	46/50			25/34	31/33	58/93 = 62	77/83 = 93			77/83 = 93	.58	G	
253					F	8/8 = 0	4/4 = 44	0		5/36	9/9	22/48	30/32	5/14	14/14	32/110 = 29	53/55 = 96			53/55 = 96	.28	G	
259					F	0/5 = 0	0/1 = 0	0		15/19	17/23			4/4	3/6	19/23 = 83	29/29 = 69			29/29 = 69	.57	F	
264					G	7/20 = 35	7/12 = 58	.20		4/7	8/12	9/12	14/18	8/9	1/3	21/28 = 75	33/43 = 77			33/43 = 77	.58	VG	
269					F	1/5 = 20	1/2 = 50	.10		9/3	1/4	2/2	4/9	1/2	2/3	3/7 = 43	7/6 = 27			7/6 = 27	.12	G	
276					G	6/6 = 100	12/12 = 100	1.0		4/4	17/20	1/1	7/9	6/6	7/21	11/11 = 100	31/50 = 62			31/50 = 62	.62	F	



PROFILE No.	USER STATEMENT CATEGORIES						PREDICTION of PERFORMANCE	ACTUAL PERFORMANCE										USER RATING									
	GOOD		FAIR		POOR			TEST COLLECTION			EST 001-006			EST 005-006			EST 001-004			OVERALL PERFORMANCE							
	1a	1b	2a	2b	1a	1b		2a	2b	RECALL %	PRECISION %	PERFORMANCE FIGURE	R	P	R	P	R		P	R	P	R	P	RECALL: PRECISION: PERFORMANCE FIGURE			
454	•						G	$\frac{2}{3} = 67$	$\frac{8}{16} = 50$	.34	$\frac{5}{12}$	$\frac{12}{25}$	$\frac{5}{11}$	$\frac{8}{13}$	$\frac{4}{6}$	$\frac{9}{8}$	$\frac{26}{46}$	$\frac{29}{46}$	$\frac{26}{46}$	$\frac{29}{46}$	$\frac{26}{46}$	$\frac{26}{46}$	$\frac{26}{46} = 57$	.27	F		
460	•		•				P				$\frac{3}{6}$	$\frac{5}{20}$	$\frac{2}{4}$	$\frac{2}{4}$	$\frac{1}{6}$	$\frac{4}{7}$	$\frac{6}{16}$	$\frac{6}{16}$	$\frac{6}{16}$	$\frac{6}{16}$	$\frac{6}{16}$	$\frac{6}{16}$	$\frac{6}{16} = 27$	.12	P		
467	•			•			F	$\frac{7}{7} = 100$	$\frac{11}{26} = 42$	.42	$\frac{4}{14}$	$\frac{10}{23}$	$\frac{4}{10}$	$\frac{4}{20}$	$\frac{2}{4}$	$\frac{8}{13}$	$\frac{22}{38}$	$\frac{22}{38}$	$\frac{22}{38}$	$\frac{22}{38}$	$\frac{22}{38}$	$\frac{22}{38}$	$\frac{22}{38} = 39$	.14	P		
472																											
479	•						G	$\frac{5}{12} = 42$	$\frac{7}{11} = 64$	.27	$\frac{10}{14}$	$\frac{21}{25}$	$\frac{2}{2}$		$\frac{7}{12}$	$\frac{18}{32}$	$\frac{7}{12}$	$\frac{18}{32}$	$\frac{18}{32}$	$\frac{18}{32}$	$\frac{18}{32}$	$\frac{18}{32}$	$\frac{18}{32} = 58$	.48	-		
483							G																				
490		•					P	$\frac{6}{7} = 86$	$\frac{6}{22} = 27$	.23	$\frac{13}{33}$	$\frac{15}{35}$			$\frac{13}{33}$	$\frac{15}{35}$	$\frac{13}{33}$	$\frac{15}{35}$	$\frac{13}{33}$	$\frac{15}{35}$	$\frac{13}{33}$	$\frac{15}{35}$	$\frac{13}{33} = 39$	.43	.17	G	
502					•		F	$\frac{9}{12} = 0$	$\frac{4}{17} = 24$	0	$\frac{5}{16}$	$\frac{12}{20}$	$\frac{10}{22}$	$\frac{14}{19}$	$\frac{8}{16}$	$\frac{11}{14}$	$\frac{23}{54}$	$\frac{23}{54}$	$\frac{23}{54}$	$\frac{23}{54}$	$\frac{23}{54}$	$\frac{23}{54}$	$\frac{23}{54} = 43$	.69	.30	G	
508					•		P	$\frac{11}{15} = 74$	$\frac{18}{45} = 40$	.30	$\frac{16}{15}$	$\frac{31}{36}$	$\frac{8}{11}$	$\frac{14}{23}$	$\frac{29}{41}$	$\frac{37}{38}$	$\frac{53}{87}$	$\frac{53}{87}$	$\frac{53}{87}$	$\frac{53}{87}$	$\frac{53}{87}$	$\frac{53}{87}$	$\frac{53}{87} = 61$	.85	.51	G	
513	•				•		F	$\frac{19}{36} = 54$	$\frac{21}{34} = 68$	.37	$\frac{12}{28}$	$\frac{24}{27}$			$\frac{20}{24}$	$\frac{33}{37}$	$\frac{33}{37}$	$\frac{33}{37}$	$\frac{33}{37}$	$\frac{33}{37}$	$\frac{33}{37}$	$\frac{33}{37}$	$\frac{33}{37} = 62$	.55	.55	G	
518		•					G	$\frac{10}{16} = 63$	$\frac{10}{10} = 100$	.63	$\frac{32}{35}$	$\frac{33}{34}$			$\frac{1}{2}$	$\frac{2}{2}$	$\frac{33}{37}$	$\frac{33}{37}$	$\frac{33}{37}$	$\frac{33}{37}$	$\frac{33}{37}$	$\frac{33}{37}$	$\frac{33}{37} = 87$	.97	.86	G	
524							F	$\frac{12}{28} = 43$	$\frac{24}{35} = 69$	.30	$\frac{57}{77}$	$\frac{67}{74}$			$\frac{21}{26}$	$\frac{25}{41}$	$\frac{78}{103}$	$\frac{78}{103}$	$\frac{78}{103}$	$\frac{78}{103}$	$\frac{78}{103}$	$\frac{78}{103}$	$\frac{78}{103} = 76$	.80	.61	G	
530					•		P	$\frac{9}{9} = 0$	$\frac{0}{31} = 0$	0	$\frac{9}{15}$	$\frac{15}{24}$			$\frac{0}{2}$	$\frac{0}{8}$	$\frac{5}{17}$	$\frac{5}{17}$	$\frac{5}{17}$	$\frac{5}{17}$	$\frac{5}{17}$	$\frac{5}{17}$	$\frac{5}{17} = 35$	.47	.16	G	
534	•						G	$\frac{2}{8} = 25$	$\frac{3}{14} = 22$	.06	$\frac{2}{7}$	$\frac{7}{11}$	$\frac{3}{5}$	$\frac{4}{8}$	$\frac{1}{4}$	$\frac{5}{5}$	$\frac{9}{19}$	$\frac{9}{19}$	$\frac{9}{19}$	$\frac{9}{19}$	$\frac{9}{19}$	$\frac{9}{19}$	$\frac{9}{19} = 47$	.44	.21	-	
539	•				•		F	$\frac{0}{0} = 0$	$\frac{0}{3} = 0$	0	$\frac{2}{2}$	$\frac{2}{2}$	$\frac{1}{1}$	$\frac{1}{2}$	$\frac{0}{0}$	$\frac{0}{0}$	$\frac{3}{3}$	$\frac{3}{3}$	$\frac{3}{3}$	$\frac{3}{3}$	$\frac{3}{3}$	$\frac{3}{3}$	$\frac{3}{3} = 100$	.75	.75	-	
544	•						G	$\frac{2}{2} = 29$	$\frac{2}{2} = 100$	.29	$\frac{9}{9}$	$\frac{2}{12}$	$\frac{5}{6}$	$\frac{3}{6}$	$\frac{3}{8}$	$\frac{3}{5}$	$\frac{3}{5}$	$\frac{9}{100}$	$\frac{9}{100}$	$\frac{9}{100}$	$\frac{9}{100}$	$\frac{9}{100}$	$\frac{9}{100}$	$\frac{9}{100} = 43$	.16	.16	G
552	•				•		G	$\frac{9}{10} = 90$	$\frac{17}{29} = 59$	.53	$\frac{9}{18}$	$\frac{18}{20}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{4}{8}$	$\frac{7}{8}$	$\frac{10}{11}$	$\frac{10}{11}$	$\frac{10}{11}$	$\frac{10}{11}$	$\frac{10}{11}$	$\frac{10}{11}$	$\frac{10}{11} = 36$	.82	.30	F	

TABLE 2. Predicted Performance, Actual Performance and User Rating

①

PROFILE No.	PREDICTION of PERFORMANCE	ACTUAL PERFORMANCE FIGURE		USER RATING
		TEST COLLECTION	201-204, 205-206, 214	
276	G	1.0	.62	F
031	G	.90	.35	G
518	G	.63	.86	G
107	G	.60	0	-
137	G	.60	.46	-
126	G	.56	.41	F
552	G	.53	.30	F
325	G	.49	.31	F/G
204	G	.48	.24	VG
391	G	.46	.55	G
194	F	.45	.41	G
082	F	.43	.21	F
220	G	.43	.46	G
435	G	.42	.84	VG
467	F	.42	.14	P
402	G	.41	.60	G
513	F	.37	.55	G
442	G	.36	.55	F
447	F	.36	.41	P
013	F	.35	.66	VG
132	G	.35	.65	G
407	P	.35	.50	F
413	P	.35	.52	VG
073	G	.34	.55	VG
454	G	.34	.27	F
022	G	.33	.66	F
165	F	.33	.71	G
052	P	.32	.35	-
036	G	.30	.45	G
508	P	.30	.51	G
524	F	.30	.61	G
544	G	.29	.16	G
097	P	.27	.32	G
314	G	.27	.50	-
479	G	.27	.66	F
170	F	.26	.18	F
043	F	.25	.24	G
490	P	.23	.17	G
209	F	.22	.61	G

PROFILE No.	PREDICTION OF PERFORMANCE	ACTUAL PERFORMANCE FIGURE		USER RATING
		TEST COLLECTION	001-002, 005-006, 014	
246	F	.21	.58	G
065	P	.20	.16	G
113	P	.20	.25	P
199	G	.20	.45	VG
264	G	.20	.58	VG
303	G	.19	.43	F
418	P	.17	.30	G
427	P	.11	.43	G
269	F	.10	.12	G
119	P	.08	.54	-
175	F	.08	.39	G
534	G	.06	.21	-
281	F	.04	.30	P
001	G	.03	.60	G
342	P	.03	.22	F
006	P	0	.44	F
056	G	0	.29	P
215	F	0	.17	F
159	F	0	.50	G
230	P	0	.59	G
253	F	0	.28	G
259	F	0	.57	F
297	F	0	.44	G
308	F	0	.33	G
319	P	0	0	P
397	F	0	.32	F
502	F	0	.30	G
530	P	0	.16	G
539	F	0	.75	-
089	F	-	.22	P
102	G	-	.17	F
145	P	-	.49	P
152	-	-	-	G
236	G	-	.40	G
242	-	-	1.0	G
287	-	-	1.0	VG
332	G	?	.25	P
337	-	-	.32	F
460	P	-	.12	P
472	-	-	.48	-
483	-	-	.46	G

**FIGURE 3. COMPARISON OF PREDICTED PERFORMANCE WITH USER ASSESSMENT IN FIRST QUESTIONNAIRE**

(SAMPLE OF 48 USERS)

<b>USER ASSESSMENT</b> <b>PREDICTED PERFORMANCE</b>	<b>POOR</b> <b>(Including very poor)</b>	<b>FAIR</b>	<b>GOOD</b>	<b>VERY GOOD</b>
<b>Poor</b>	3 (6%)	1 (2%)	7 (15%)	
<b>Fair</b>	1 (2%)	3 (6%)	12 (15%)	1 (2%)
<b>Good</b>		8 (17%)	7 (15%)	5 (10%)

**FIGURE 4. COMPARISON OF PREDICTED PERFORMANCE WITH USER ASSESSMENT IN SECOND QUESTIONNAIRE (SAMPLE OF 48 USERS)**

<b>USER ASSESSMENT</b> <b>PREDICTED PERFORMANCE</b>	<b>POOR</b> <b>(Including very poor)</b>	<b>FAIR</b>	<b>GOOD</b>	<b>VERY GOOD</b>
<b>Poor</b>	1 (2%)	3 (6%)	5 (10%)	2 (4%)
<b>Fair</b>		3 (6%)	12 (25%)	2 (4%)
<b>Good</b>		2 (4%)	13 (27%)	5 (10%)

**FIGURE 2. PERFORMANCE FIGURES FOR PROFILES DERIVED FROM STATEMENTS GRADED POOR, FAIR AND GOOD**

PREDICTION	NO. OF PROFILES	MEAN ANNUAL OVERALL PRECISION	
		R1	R1/2
Poor	17	27	66
Fair	26	29	63
Good	31	34	75

**TABLE 3. AVERAGE PERFORMANCE OF PROFILES IN VARIOUS PREDICTED PERFORMANCE CATEGORIES**

PERFORMANCE FIGURE	PERFORMANCE CATEGORY				
	VP	P	F	G	VG
Average Test Collection performance figure for Predicted Performance	-	$\frac{2.61}{12} = .22$	$\frac{4.17}{15} = .28$	$\frac{11.04}{27} = .41$	-
Average Experimental Service performance figure for User Ratings	-	$\frac{2.47}{9} = .27$	$\frac{7.31}{19} = .38$	$\frac{15.56}{35} = .44$ $\left. \begin{array}{l} \text{---} \\ \text{---} \end{array} \right\} .47$	$\frac{4.84}{8} = .61$
Average Experimental Service performance figure for Predicted Performance	-	$\frac{6.07}{17} = .36$	$\frac{10.0}{25} = .40$	$\frac{13.83}{30} = .46$	-

TABLE 4. OVERALL PRECISION REFERENCE FOR WEEKS 011-056  
and USER SATISFACTION ASSESSMENT IN QUESTIONNAIRE  
FOR SAMPLE SET OF USERS

PROFILE NO.	ASSESS- MENT at QUESTION- NAIRE 2 2	OVERALL PRECISION FOR YEAR		PROFILE NO.	ASSESS- MENT at QUESTION- NAIRE 2 2	OVERALL PRECISION FOR YEAR	
		R1	R1/2			R1	R1/2
001		35	92	269	G	39	75
006		13	63	276	VG	52	90
013	VG	39	81	281		20	40
022	G	25	86	287	VG	59	95
031	G	68	89	297	G	19	72
036	G	77	99	303	G	22	86
043	G	43	72	308	G	56	88
052	G	28	68	314	G	16	60
056		48	54	319	P	10	67
065	G	22	74	325	G	10	73
073	VG	58	80	332	F	17	64
082		9	36	339	G	12	37
089		21	42	342	F	26	78
097	VG	48	77	392	G	16	60
102		0	29	397	G	37	67
107		7	58	402	VG	52	78
113		15	99	407		44	68
119		38	75	413		52	82
136	G	28	83	418	VG	13	41
132		40	60	427		19	52
137	G	50	82	435	VG	53	88
145	F	28	77	442	G	39	75
152	G	15	75	447		19	57
157	G	23	64	454		15	74
165	G	67	94	460	F	5	23
170	F	13	37	467	F	13	49
175	G	22	68	472	G	21	81
194	G	32	76	479	F	1	55
199	VG	35	96	483	G	25	65
204	G	27	78	490	G	75	98

TABLE 4 (continued)

PROFILE NO.	ASSESSMENT QUESTIONNAIRE 2	OVERALL PRECISION FOR YEAR		PROFILE NO.	ASSESSMENT QUESTIONNAIRE 2	OVERALL PRECISION FOR YEAR	
		R1	R1/2			R1	R1/2
209	G	8	67	502	F	7	44
215	G	13	40	508	G	11	51
220	G/VG	38	79	513	G	30	59
230	G	34	72	518	G	78	93
236		36	89	524		57	82
242	F	36	52	530	G	18	28
246	VG	67	97	534	F	15	54
253		26	84	549		19	42
259		4	30	544	G	19	69
264	G	54	83	552	F	19	80

2. Predicted performance and user assessment at first and second Questionnaires.

FILE	PREDICTION	Q1	Q2	PROFILE	PREDICTION	Q1	Q2
	F	VG	VG	297	F	G	G
	G	F	G	303	G	F	G
	G	G	G	308	F	G	G
	G	G	G				
	F	G	G	319	P	P	P
	P	G	G	325	G	F/G**	G
	G	VG	VG	342	P	F	F
	P	G	VG	391	G	G	G
	G	F	G	397	F	F	G
	P	P	F	402	G	G	VG
	F	G	G	418	P	G	VG
	F	G	G				
	F	F	F	435	G	VG	VG
	F	G	G	442	G	F	G
	F	G	G	460	P	P	F
	G	VG	VG	467	F	P	F
	G	VG	G	479	G	F	F
	F	G	G	490	P	G	G
	F	F	G	502	F	G	F
	G	G	G/VG*	508	P	G	G
	P	G	G	513	F	G	G
	F	G	VG	518	G	G	G
	G	VG	G	530	P	G	G
	F	G	G	544	G	G	G
	G	F	VG	552	G	F	F

\* Taken as G

\*\* Taken as F

## Chapter 23

### EFFECT OF PROFILE ANALYSIS AND MODIFICATIONS ON USER'S SATISFACTION WITH SERVICE

It might be assumed that time and effort spent on analysis and modification of profiles would result in an improvement in performance.

An examination was made to discover the extent to which this was true.

#### Method

Using the data presented in Appendix 9B, a table (Figure 1.) was compiled categorizing users according to their assessment of the service in the first and second questionnaires. For each category the average time spent per profile in analysis and modification was worked out.

Figure 2 shows the improvement for all profiles comparing the first questionnaire assessment with that for the second questionnaire. Thus an improvement of +1 includes all profiles improving from Good to Very Good, Fair to Good, Poor to Fair and Very Poor to Poor. Users who gave a half way assessment e.g. F/G are excluded. The net improvement is plotted against the average time spent per profile in analysis and modification.

Since only one profile is represented in the figures for +3, -2 and -1 the points for these on the graph should be discounted. The remaining points cannot be claimed to show a pattern.

In figures 3, 4 and 5 the figures for each profile are shown in six categories according to whether the assessment in the first questionnaire was Very Poor, Poor, Fair, Fairly Good, Good or Very Good.

Again no curve could reasonably be drawn.

### Conclusion

From an examination of the results of this study there seems no reason to believe that there is any direct relationship between the amount of time spent in profile analysis and modification and improvement in user satisfaction. There are perhaps several reasons why this might be so.

- 1) It is difficult to determine how much of the time spent is concerned actually in improving the profile and how much in collection and analysis of data to decide what modification is needed. For complete profiles the time spent in analysing large numbers of relevance assessments can be considerable and the result of much work may be a very minor change to the profile.
- 2) Inherently unsatisfactory profiles may take up considerable analysis and modification time without the basic fault being removed.

1st Q	2nd Q	No. of profiles	Total time in mins	Average time per profile in mins
VG	VG	12	935	77.9
VG	G	6	535	89.2
G	VG	20	1412	70.6
G	G	89	5505	61.9
G	F	19	1835	96.6
G	F	1	175	175.0
G	VP	1	90	90.0
F	VG	13	1395	107.3
F	G	53	6130	115.7
F	F	29	2742	94.6
F	P	2	215	107.5
P	VG	1	190	190.0
P	G	5	485	97.0
P	F	12	1580	131.7
P	P	7	645	92.5
VP	F	3	595	198.3
VP	VP	1	25	25.0
G	G/VG	1	0	0
F/G	VG	1	25	25.0
F/G	G	2	155	77.5
F/G	F	1	0	0

Figure 1. User satisfaction with SDI Service at beginning and end of weekly service, and time spent in profile analysis and modification.

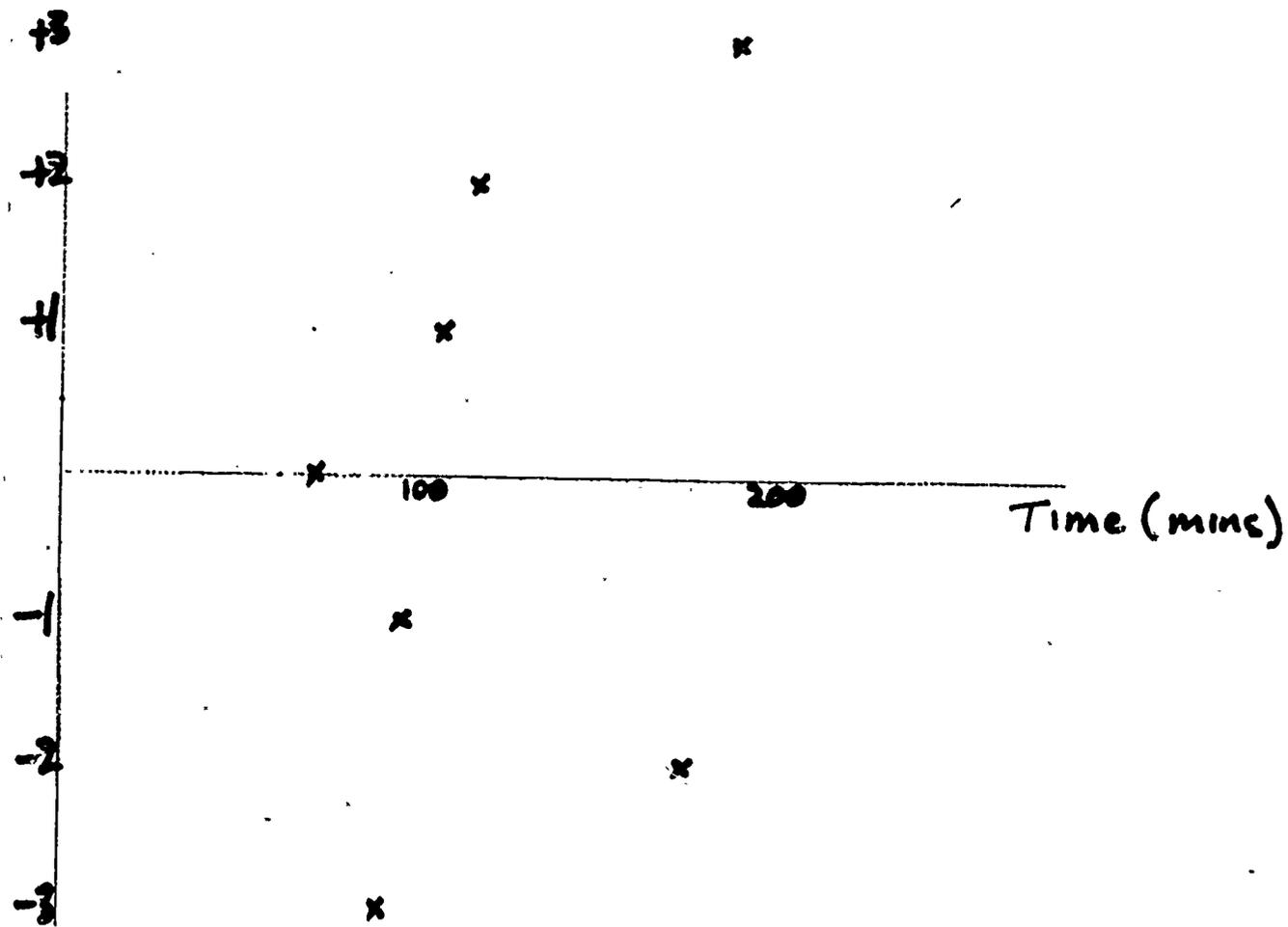
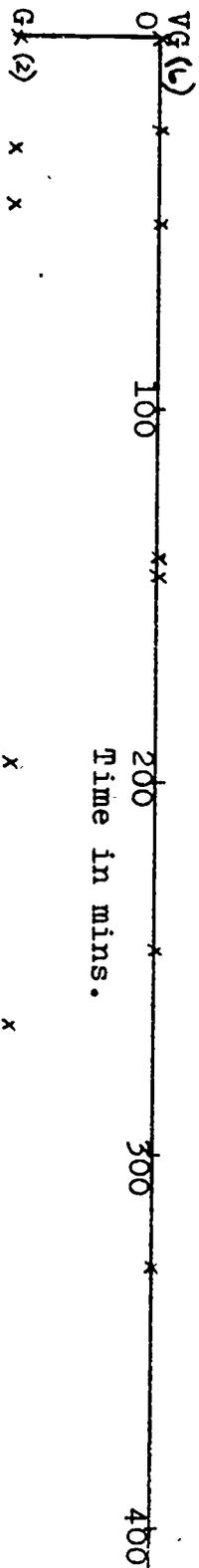


Fig. 2 Improvement in user satisfaction with service in relation to time spent in profile analysis and modification

Improvement	No. of profiles	Total time (mins)	Avg. time per profile
+ 3	1	190	190
+ 2	21	2475	118
+ 1	85	9122	107
0	138	9852	71
- 1	27	2585	96
- 2	1	175	175
- 3	1	90	90

Second Questionnaire Assessment.

1. Profiles originally assessed as VG.



2. Profiles originally assessed as G.

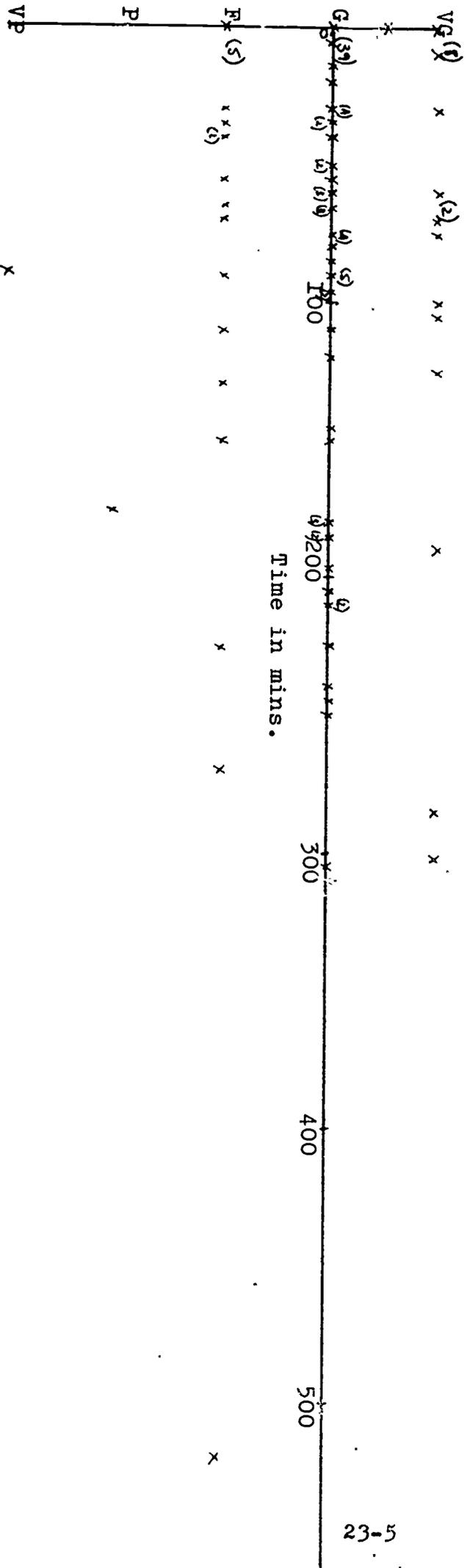
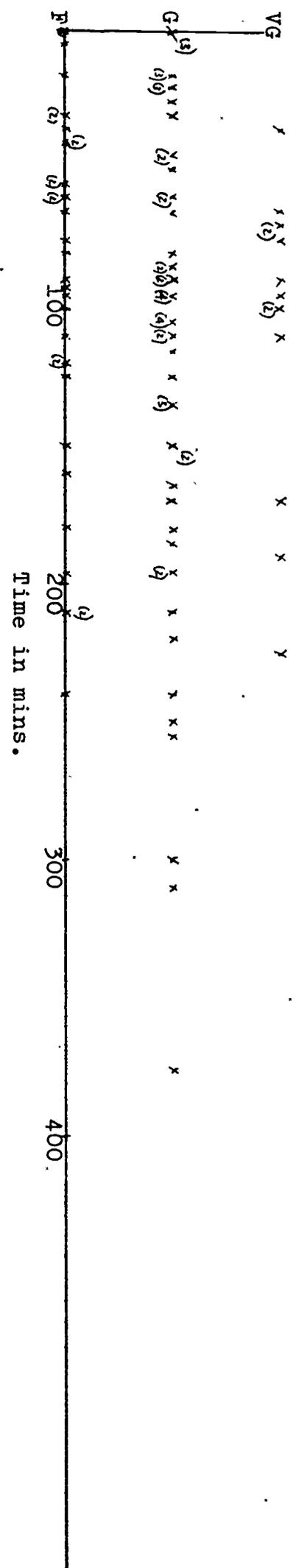


Fig. 3. User assessment of SDI service in relation to time spent in profile analysis and modification.

Second Questionnaire Assessment.

3. Profiles originally assessed as F.



4. Profiles originally assessed as F/G.

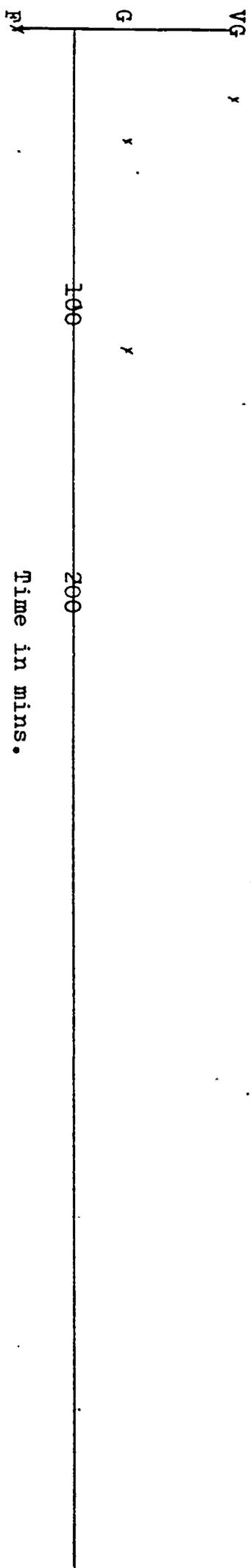
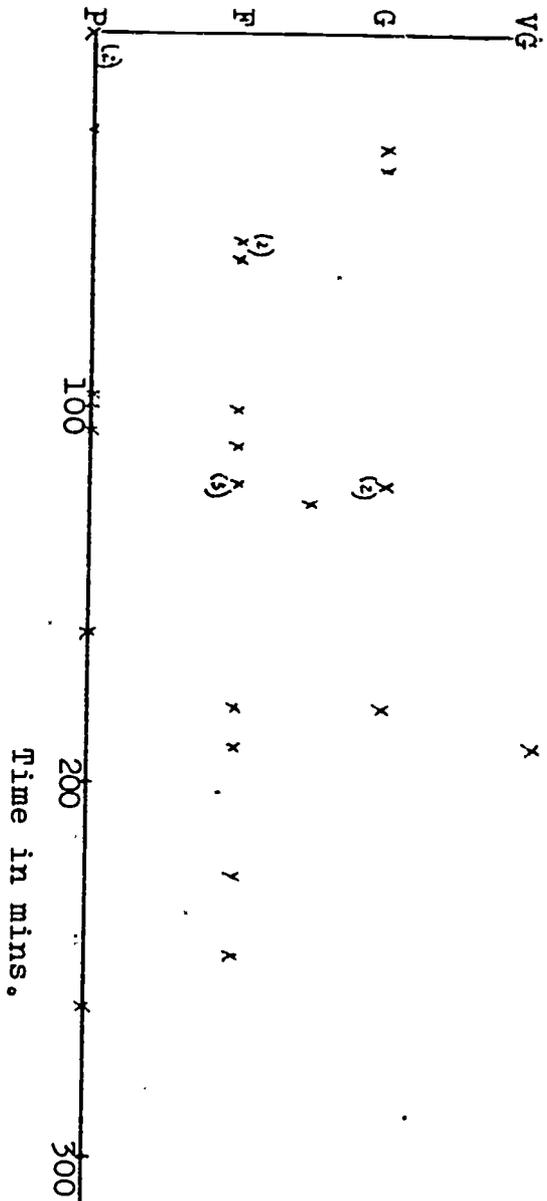


Fig. 4. User assessment of SDI service in relation to time spent in profile analysis and modification.



Second Questionnaire Assessment.

5. Profiles originally assessed as P.



6. Profiles originally assessed as VP.

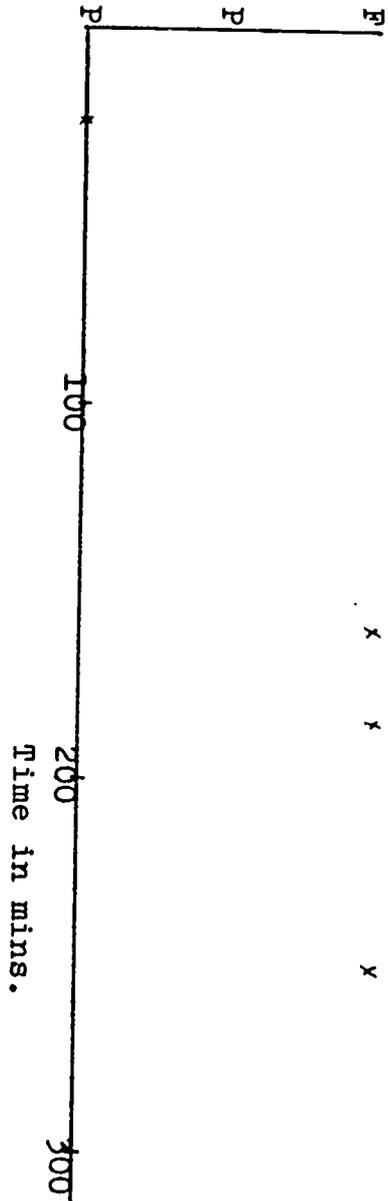


Fig. 5. User assessment of SDI service in relation to time spent in profile analysis and modification.

## Chapter 24

### EFFECT OF NUMBER OF DOCUMENT NOTIFICATIONS RECEIVED BY USER ON HIS RELEVANCE ASSESSMENT

It has been suggested that there may be a tendency for users to vary their relevance assessments depending on the number of document notifications they receive. Thus if a user receives a large amount of documents he is likely to be more selective in his assessment than if he receives a smaller number and will tend to mark fewer relevant, with a consequent tendency to give lower Precision figures.

To determine whether this very plausible hypothesis could be shown to have any basis in fact an examination was made of the data obtained during the investigation.

Figure 1 gives for each of weeks 011-050, the average number of notifications sent to users and the Precision performance  $R_{1/2}$  for both average of numbers and average of percentages. In Figure 2 the number of notifications are plotted against Precision (average of numbers) and Figure 3 shows the same information but using average of ratios instead. In neither case would there appear to be any relationship between the average number of notifications received and the Precision performance.

#### CONCLUSION

Data available from the SDI service do not appear to support the suggestion that users tend to mark a small proportion of documents relevant when they are presented with larger numbers of notifications.

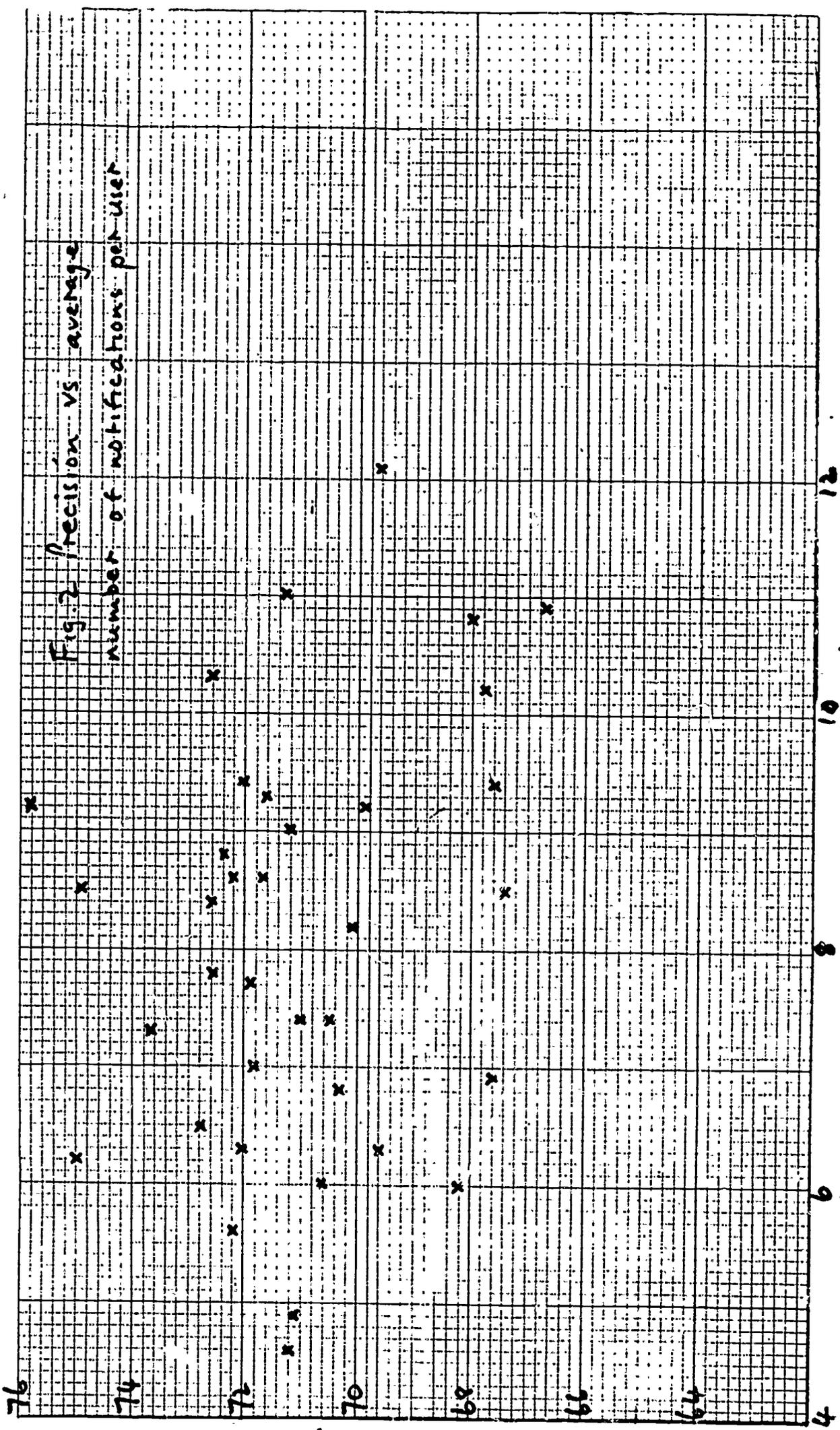
However in view of the considerable number of uncontrolled variables including, most noticeably, the completely different sample of source journals and documents represented in each week's run, it is in no way possible to draw any stronger conclusions.

FIGURE 1. AVERAGE NUMBER OF NOTIFICATIONS AND PRECISION FIGURES  
FOR WEEKS 011-050

Week No.	Average notifications per user	% Precision	
		Average of nos.	Average of %
11	7.4	70.5	69.5
12	8.8	72.4	71.5
13	10.9	66.7	66.5
14	12.1	69.6	70.1
15	10.2	67.8	66.4
16	9.4	67.6	67.2
17	10.8	68.0	64.8
18	8.5	67.4	67.5
19	9.4	72.0	69.6
20	9.3	71.8	68.0
21	8.6	71.7	69.4
22	8.2	70.1	68.7
23	9.2	69.5	67.7
24	6.9	67.6	67.9
25	6.3	72.0	71.1
26	10.8	70.1	68.4
27	8.6	71.5	70.1
28	5.6	72.2	72.1
29	4.9	71.1	70.7
30	6.0	70.3	68.1
31	7.7	71.9	71.0
32	9.2	75.8	74.9
33	11.0	71.3	71.5
34	9.0	71.2	69.3
35	9.5	74.4	72.0
36	7.4	71.0	68.6

Figure 1. (Contd)

Week No.	Average notifications per user	% Precision	
		Average of Nos.	Average of %
37	8.5	74.9	71.8
38	8.4	72.6	72.5
39	8.6	72.2	71.8
40	10.3	72.6	71.9
41	9.0	70.9	69.6
42	6.3	69.6	67.0
43	4.6	71.2	69.3
44	6.5	72.8	69.8
45	7.8	72.6	68.2
46	7.3	73.7	72.6
47	6.8	70.3	67.5
48	6.2	75.0	72.9
49	6.0	68.2	64.0
50	7.0	71.8	69.9



Precision (average of numbers)

Average number of notifications

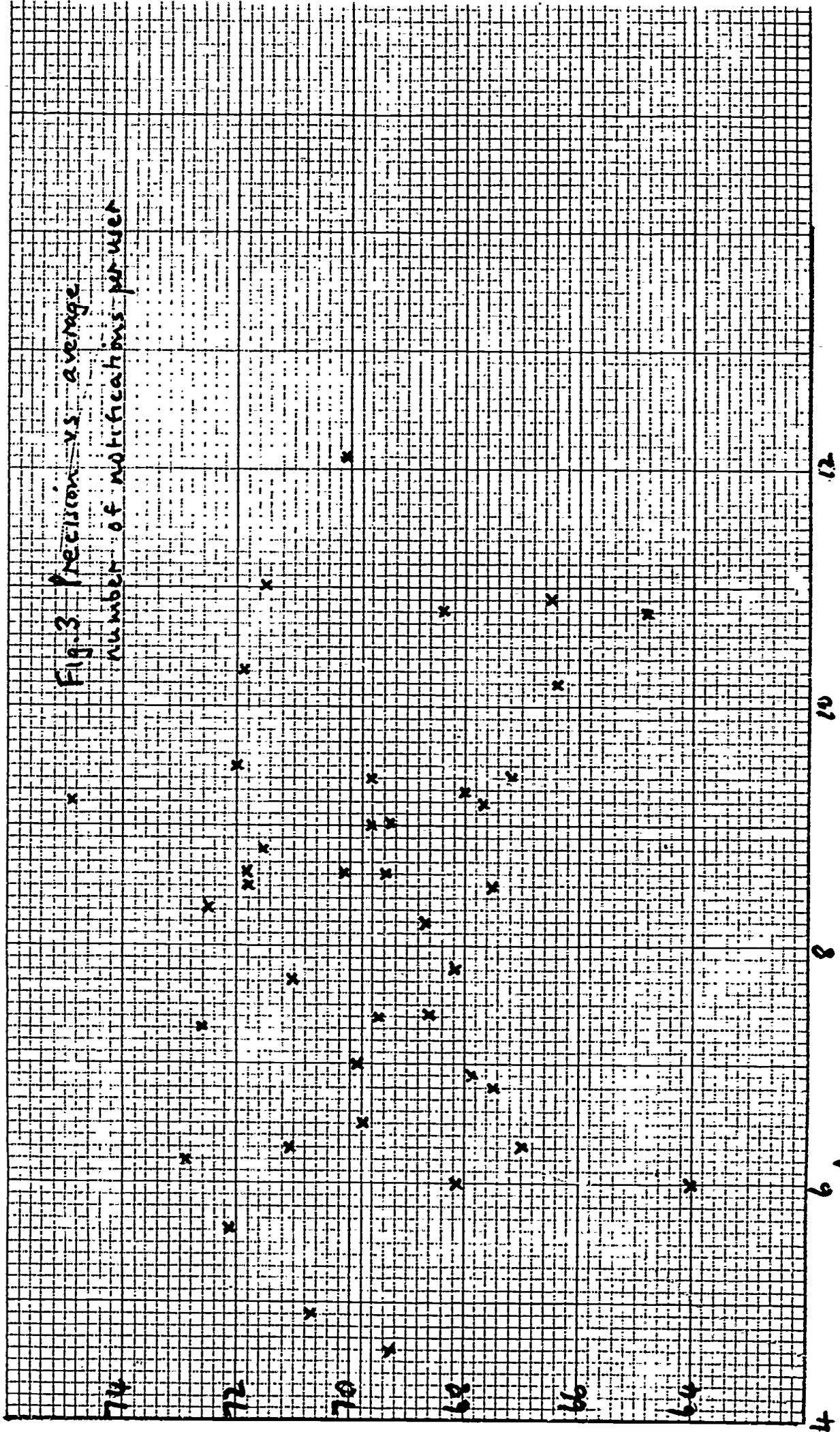


Fig. 3 Precision vs. average number of notifications per user

Precision (average of %)  
24-5

Average number of notifications

## Chapter 25

### COMPARISON OF RELEVANCE ASSESSMENTS BASED ON THE FULL DOCUMENT WITH THOSE BASED ON CARD NOTIFICATIONS GIVING VARYING AMOUNTS OF DOCUMENT INFORMATION

In the vast majority of SDI systems the user receives as the product of the search, not the selected documents themselves, but some substitute (surrogate), commonly including the author, title, source reference, index terms and possibly an abstract of the document. The object is to give him sufficient information in a convenient form to allow him to decide whether the document is relevant and to choose whether he wishes to obtain the full text.

The question arises as to the information which should be provided in the document surrogate to allow the user to make a reliable judgment of the relevance of the original to his needs.

#### METHOD

Four different types of notification cards were produced containing different amounts of information. The basic information was in each case the same i.e. author, title of document and journal reference. In the case of Group II notifications this was the only information supplied. Groups I, III and IV contained additional information as follows:-

- Group I    Basic information plus index terms assigned by the SDI indexers. (This was the information normally supplied on SDI notification cards).
- Group II    Basic information only
- Group III    Basic information plus abstract.
- Group IV    Basic information but with title augmented to make it more explanatory where necessary.

Normally all documents received were numbered serially after indexing but for this study they were numbered before being passed to the indexers. The numbering was carried out as follows. The first four documents were numbered 062001 - 004, the second four 062011-014 etc. Thus each document was assigned to one of the four classes depending on the terminal digit. This ensured that documents arriving together e.g. articles in the same journal were not all grouped together and gave a reasonably comparable sample in each group, while allowing documents to be numbered and processed immediately on receipt as was essential for the currency of the SDI Service. At the same time the coincidence of the Group number and the terminal digit simplified considerably the later procedures and analysis.

After numbering, the documents were indexed as usual and in addition an augmented title was written where necessary for documents assigned to Group IV (terminal digit 4), and for those documents in Group III (terminal digit 3) which did not have a usable abstract, an abstract was written. The size of abstract was limited by the space available on the card. A maximum of approximately 80 words was possible and some abstracts had to be shortened.

All the data except the abstract was input to the computer by means of punched paper tape as normal, and thus appeared on the notification cards produced from the tapes via offset litho masters. All that was necessary was to add the abstracts for the Group III documents to the offset litho master. Thus all users received a set of notifications containing a mixed set of Group I, II, III or IV notification cards. No note explaining the presence of abstracts on some cards or the absence of index terms on others was sent to users and it is interesting that very little comment was received from them.

Examples of the various notifications are shown in Figures 1 - 4.

Several weeks later all users who returned relevance assessments for these documents were sent copies of the documents and asked for relevance assessments based on the full document. To save expense only approximately half of the documents were treated in this way. The sample was made by sending copies of only those documents having an even fifth digit in the serial number e.g. 062362 thus giving a sample evenly spread throughout the collection and therefore not confined to articles from a few journals. The covering letter sent with the sample of documents is shown in Appendix 10.

For each document assessed, the following information was recorded - user number, document number, relevance assessment based on card notification, and relevance assessment based on full document. The relevance assessments were in each case either 1, 2 or X, being respectively 'highly relevant', 'partially relevant' and 'not-relevant'.

## RESULTS

The results are shown in figure 5

It is clear from the figures that the best correlation with the full document assessment is given by the assessment based on the notification which include the augmented title and by those which include an abstract. The worst agreement is that for the 'title only' notifications, though the range between all four is not very great. Perhaps suprising is the fact that provision of an abstract does not appear to increas the chance of agreement between the card assessment and the full document assessment, as compared with the augmented title.

It may be concluded that an abstract does not make a significant contribution in helping the user to decide relevance of the original document.

**FIGURE 1.** Group I notification card - normal SDI notification

062511; 99;

Upper-atmosphere winds and their interpretation-II  
Turbulence in the lower E-region;

D. Layzer and J.F. Bedinger;

Planet Space Sci. Vol.17 No. 11 1891-1911 November  
1969;

Ionosphere;  
Variations;

Atmospheric Currents;  
E. Region;

**FIGURE 2.** Group II notification card - title only

062012; 99;

Using thermostimulated exoelectron emission for  
ionizing radiation dosimetry;

A.I. Beskorskii and others;

Instrum. Exper. Tech. No. 1 35-38 January-February  
1969;

FIGURE 3. Group III notification card - abstract

062533; 99;

Measuring narrow f.m. deviation;

R.A. Kennedy;

Marconi Instrum. Vol.12 No.3 54-55 September  
1969;

On the F.M./A.M. Modulation Meter type TF 2300 the lowest deviation range is 5 KHz full-scale. Although this is adequate for conventional applications it does not provide sufficient discrimination to measure the very low spurious f.m. from mobile or broadcast transmitters. This requirement can be met by using an external voltmeter to measure the l.f. output of the Modulation Meter which enables deviations as low as 10 Hz to be measured.

FIGURE 4. Group IV notification card - augmented title

062554; 99;

An inexpensive multichannel scaler with channel widths of less than 1 microsec: uses a time-to-pulse-height converter;

E.C. Silverberg;

Rev. Sci. Instrum. Vol.40 No.11 1530-1504  
November 1969;

FIGURE 5. Relevance assessments of various card notifications compared with those based on full document.

GROUP	Assessed higher on card		Assessed higher on document		Equal assess-ment		TOTAL
	No.	%	No.	%	No.	%	
I	45	19	31	13	160	68	236
II	53	18	50	17	195	65	298
III	37	18	25	12	148	71	210
IV	29	13	34	15	162	72	225
TOTAL	164		140		665		969

## Chapter 26

### Effect of including author names as search terms in SDI Profiles.

The statement of Information Requirements form sent to each user asked for details of the subjects on which he would wish to receive information. No attempt was made to discover the names of authors whose work would be likely to be of interest. Thus except for one or two users who volunteered some information on this point, SDI profiles throughout the Investigation did not contain author names and searched solely on subject matter.

It was however intended to ask users for names of suitable authors and to include these in the profiles at some convenient time during the Investigation. For this reason the authors of documents were included as searchable elements in the document record during the early stages of the Investigation. (It should perhaps be pointed out that to allow a search on authors to be made the names had to be included as descriptors along with the subject descriptors in the document record). Until the author names could be included in profiles this part of the record was not usable and when a way was being sought of reducing the punching load from week 026, the author descriptors were obvious candidates for dropping until the temporary flood of input had been dealt with. The inflow of documents however remained at a high level for a long time and it was not possible to consider reinstating the author elements of the record even though lists of authors had now been obtained from the users for inclusion in profiles. It will be realised too that the punching effort required to add an average of say 20 author names to each of some 600 profiles was in itself not readily available at this time. Apart from the question of punching effort it must be remembered that owing to the computer program difficulties, profile modification was at no time a straightforward matter and there was a natural reluctance to face the disruption of the service that could result from profile modification on this scale.

Nevertheless the facility of searching by author names is important and it was not satisfactory to carry out the Investigation without at least examining the effect of including this option. A small investigation was therefore planned to discover how the inclusion of author names in SDI profiles would affect the performance of the SDI service.

### Methods

Given the constraints mentioned above i.e. the need to keep the punching load to a minimum and to disturb the profiles as little as possible the following methods were adopted.

- 1) The author descriptors were added to the documents as surname without initials e.g. FREEMAN rather than FREEMAN F.J. The difference may appear slight but in fact reduced to less than half the time taken and the error rate in punching since errors occur mainly with the spacing and punctuation of initials.
- 2) Instead of adding the authors to the individual profiles, a relatively small number (25) of special 'author' profiles were constructed made up solely of the author names for the profiles as a whole. In these again the surnames only were used. The use of these special profiles avoided disturbance of the individual profiles and kept the matching output for the 'authors investigation' separate from the normal service output and available for separate scrutiny.

Thus the function of these author profiles was to select articles by authors with given surnames. These had then to be sorted manually to obtain those by the wanted authors i.e. those with given initials. The particular users interested in these authors could then be looked up in a card index and the appropriate notifications sent to them for relevance assessment.

This was, of course, a laborious method but it was considered better to accept it for the limited duration of the study rather than abandon the project as impossible. In the event however the extra work required to overcome the deficiencies in the computer service was so enormous that the study would not have been attempted if this could have been foreseen.

### Organisation

A letter (Appendix 10A82) was sent to all users inviting them to supply details of any authors whose work was likely to be of interest.

Although a reply to the letter was left optional most participants did respond and few gave less than ten names.

It was intended to compile the author profiles and to run them for ten consecutive weeks and to use the output for this particular study. However considerable difficulties were encountered in getting the profiles on file and the period of time that each profile was on file varied.

At the end of the period the items produced by the author profiles were sent for relevance assessment to the appropriate participants.

In view of the fact that each of the author names given by a participant was on file for a different length of time the results could not be used to show what proportion of documents retrieved might be caused by author names in normal circumstances. A manual search was therefore undertaken to assess this.

The output for each of 24 randomly-selected profiles was examined over a period of ten weeks to discover whether any of the documents were written by authors named. The results of this are shown in Table 1.

### Results

1. Of the participants who supplied names of authors, 93 were sent 259 documents to assess. These comprised 147 individual documents.
2. Relevance assessments were received for 163 of the 259 documents, 93 being assessed as relevance 2.
3. 80 (49%) of the 163 documents were also retrieved by the subject profiles i.e. 83 (51%) were retrieved only by author. Of these, 61 (74%) were classed as relevant (R1 or R2) and 34 (40%) as highly relevant (R1).
4. Of the 80 documents retrieved by both subject and author, 79 (99%) were relevant and 62 (78%) were highly relevant.
5. From the results of the manual matching (Table 1) it can be seen that of 971 relevant documents selected by the subject profiles of 24 participants only 29 (3%) would also have been selected by author names.

It must be remembered that some of the names supplied by participants were those of foreign authors which could not be expected to retrieve documents from largely English-language periodicals.

However the number of documents retrieved by subject from a file including foreign language articles would have been proportionately larger and there is no reason to believe that the percentage of documents retrieved by author names would have been greater.

### Conclusions

One is hesitant to draw very firm conclusions from the study. However, it seems clear that with such a small percentage of relevant documents retrieved by author compared with those retrieved by subject, the overall performance of the SDI service is unlikely to have been affected seriously by the exclusion of author names from profiles.

Although the percentage of documents retrieved by author is small the relevance of such documents on average is much higher than those retrieved by subject i.e. 86 per cent are relevant and 57 per cent highly relevant compared with figures of approximately 70 per cent and 33 per cent for documents retrieved by subject.

Table 1

Profile No.	No. of RD1 & RD2 notifications sent for weeks 050 - 059											Total	No. of author matches for 050 - 059
	050	051	052	053	054	055	056	057	058	059			
012	1	2	2	2	4	4	-	1	-	-	16	1	
034	14	8	8	16	11	12	13	5	19	17	122	-	
072	3	6	4	5	2	7	1	1	1	4	34	2	
079	-	-	1	-	-	1	-	-	1	-	3	-	
113	3	3	4	3	1	7	7	7	6	2	43	-	
141	12	6	5	21	6	7	3	3	8	7	78	-	
158	2	1	7	-	3	9	7	-	5	-	34	-	
165	4	11	12	9	12	6	3	9	5	10	81	1	
177	1	2	-	4	1	-	-	6	2	4	20	3	
17	-	1	-	-	3	-	-	-	1	10	15	1	
244	-	1	-	-	-	1	-	-	1	-	3	1	
287	4	2	-	-	3	4	8	3	3	2	29	3	
321	-	6	5	4	7	8	1	7	4	6	48	3	
354	-	-	-	-	-	-	3	-	-	-	3	1	
364	-	2	6	1	2	11	5	8	7	2	44	1	
385	3	4	7	-	5	5	2	4	3	11	44	1	
42	1	-	3	1	1	5	-	10	3	2	26	-	
511	8	4	9	2	-	12	-	-	-	8	43	-	
532	-	1	1	-	-	-	-	1	-	-	3	1	
540	3	2	4	-	2	2	1	3	1	3	21	5	
573	2	8	9	12	8	6	2	6	1	9	63	-	
588	-	-	-	-	6	10	18	7	6	8	55	1	
615	8	2	20	1	4	13	8	18	11	5	90	3	
638	6	3	1	8	1	-	8	9	8	9	53	1	
											971	29	

A dash indicates that no assessment is available either because there was no matching for that week, or because assessments were not returned. e.g. In the case of 079 where the profile was on a very limited topic the dashes indicate the former. In 588 where the output was normally quite high, from 050-053 the user failed to return assessments.

## Chapter 27

### VARIATION OF USER SATISFACTION WITH COMPILER OF PROFILE

All search profiles used in the Investigation were compiled internally by INSPEC staff. The majority of profiles were compiled by two of the Indexer/Analysts who remained on the staff throughout the Investigation. The first of these (Compiler No. 1) was responsible for 230 profiles and the second (Compiler No. 2) for 219 profiles. The remainder of the profiles, 126 in all, were compiled by various other members of the staff during the course of the investigation and, though most of them owe more to one person than another they do not form a homogeneous group since many after compilation by one person, underwent subsequent analysis and modification by others.

However for the sake of identification these profiles are assigned to compiler No. 3. Details of the profiles assigned to each compiler are given in Appendix 27 A.

It is obviously of some interest where profiles are compiled by different people to discover whether the performance of the sets of profiles differs and what factors might account for any difference found.

As far as formal education is concerned Compilers 1 and 2 may be considered generally equal in that they were both science graduates without directly applicable training in the specific field covered by the investigation i.e. electronics research, but they were specialists in the field of physics.

In terms of experience of information work, Compiler 1 had several years advantage over Compiler 2 who was a recent graduate. However it is likely that the greatest difference between the two was greatest in temperament and approach to the work, Compiler 1 adopting a systematic analytical approach while Compiler 2 took perhaps a more intuitive line.

A comparison was made of the profiles compiled by each compiler to discover whether the respective users were equally satisfied with the service received. The first service questionnaire returns for each Compiler are shown in Figure 1, and those for the second questionnaire in Figure 2.

The figures are shown as histograms for each Compiler in Figures 3-5. For the first questionnaire returns, the general distribution of user assessments is very similar, although the figures for each category of satisfaction differ in some cases e.g. 38% Fair for Compiler 1 against 33% for Compiler 2. However it is not possible on these figures to say that there is any overall difference in user satisfaction for the two compilers since the total in each case who express positive satisfaction (Good or Very Good) is very similar, 53.5% for Compiler 1 and 54.4% for Compiler 2 with a higher number of Very Good assessments in the latter case. This slightly greater degree of satisfaction for Compiler 2 profiles is offset by the fact that 12.5% of Compiler 2 users express dissatisfaction (Poor or Very Poor) as against 8.4% for Compiler 1 profiles. The Fair assessment is really too neutral to be considered as evidence on its own.

It can only be considered from these figures that there appears to be no appreciable difference between the two sets of profiles in terms of user satisfaction.

It must be admitted that this was a most unexpected result since for various reasons the Compiler 1 were thought likely to be superior in this respect.

The comparison of Compiler 1 and 2 profiles with Compiler 3 profiles is striking. The histograms show very great differences, with Compiler 3 users having the highest percentage of Very Good profiles but the lowest percentage of Combined Good and Very Good profiles. It is, however, not possible to consider this group of profiles as homogeneous in view of the various people who were responsible for them. In addition, many of these profiles were compiled at a very early stage in the investigation before the Thesaurus was complete and were therefore in some cases deficient in some terms.

In the second questionnaire it can be seen that though satisfaction appears to have increased for all three groups of profiles, the increase in satisfaction is greater for Compiler 1 profiles than for Compiler 2 profiles. In the former group the percentage of users expressing positive satisfaction (Very Good or Good) has increased by approximately 20 percent while for the latter group the increase is only approximately 13 percent.

Increased satisfaction is however greater in the case of the compiler 3 profile where positive satisfaction has increased from 48 to 72 percent i.e. by 24 percent.

To discover whether this improvement could be related to the times spent in profile analysis and modification by each compiler, the relevant data were obtained and are shown in Figures 6-8.

The average time for profile analysis and modification for each compiler is as follows:-

Compiler 1	-	84 minutes
Compiler 2	-	91 minutes
Compiler 3	-	107 minutes

It appears therefore that this in itself does not affect the increase in user satisfaction.

### CONCLUSIONS

It would be unwise to attempt firm conclusions but one or two are suggested.

- 1) User satisfaction appears to be largely determined by factors other than the person compiling the profile since overall satisfaction is very similar for both Compiler 1 and Compiler 2.
- 2) Nevertheless by subsequent analysis and modifications an increase in satisfaction is attainable and Compiler 1 appears to have achieved a somewhat greater increase than Compiler 2.
- 3) Compiler 3 profiles initially exhibit a different satisfaction distribution from the other profiles, presumably reflecting a mix of profiles compiled at widely different times by different people. However it is noticeable that by the time of the second questionnaire after revision they exhibit much the same characteristics as the other two groups of profiles. The greater increase in satisfaction may well be due to their starting from a lower level of satisfaction originally.

It is perhaps surprising that the individual profile compiler appears to make only a relatively small impact on user satisfaction. Some possible reasons may be as follows:-

- 1) Profile compilation, analysis and modification is aimed primarily at achieving good performance in terms of the performance measures used in the system.

- 2) User satisfaction, though broadly related to measured performance of the profiles does not correspond directly to it in all cases. Thus it is possible for a compiler to achieve a higher performance in terms of greatly increased Precision with relatively little loss of Recall, but give less satisfaction to a user who is concerned with high Recall almost to the exclusion of Precision.
- 3) The user's satisfaction with the performance of his profile is almost entirely determined by the extent to which he has clearly stated his requirements, how easily these can be coded, how consistently he stands by these stated requirements in assessing the service, and, not least, by his temperament. Given all these major factors in determining user satisfaction, the scope for the profile compiler may be very restricted.

FIGURE 1.

USER SATISFACTION AT FIRST QUESTIONNAIRE AND  
PROFILE COMPILER

Compiler No.	Very Poor No.	%	Poor No.	%	Fair No.	%	Good No.	%	Very Good No.	%	TOTAL
1	3	(1.8)	11	(6.4)	65	(38)	86	(50)	6	(3.5)	171
2	4	(2.5)	16	(10)	54	(33)	76	(47)	12	(7.4)	162
3	2	(2.2)	10	(11)	35	(39)	33	(37)	10	(11)	90
<b>TOTAL</b>	<b>9</b>	<b>(2.1)</b>	<b>37</b>	<b>(8.8)</b>	<b>154</b>	<b>(36)</b>	<b>195</b>	<b>(46)</b>	<b>28</b>	<b>(6.6)</b>	<b>423</b>

FIGURE 2.

USER SATISFACTION AT SECOND QUESTIONNAIRE AND  
PROFILE COMPILER

Compiler No.	Very Poor No.	%	Poor No.	%	Fair No.	%	Good No.	%	Very Good No.	%	TOTAL
1	-	-	5	(4.1)	28	(23)	67	(55)	22	(18)	122
2	1	(0.9)	3	(2.6)	35	(30)	61	(52)	17	(15)	117
3	1	(1.3)	5	(6.5)	16	(21)	39	(51)	16	(21)	77
<b>TOTAL</b>	<b>2</b>		<b>13</b>		<b>79</b>		<b>167</b>		<b>55</b>		<b>316</b>

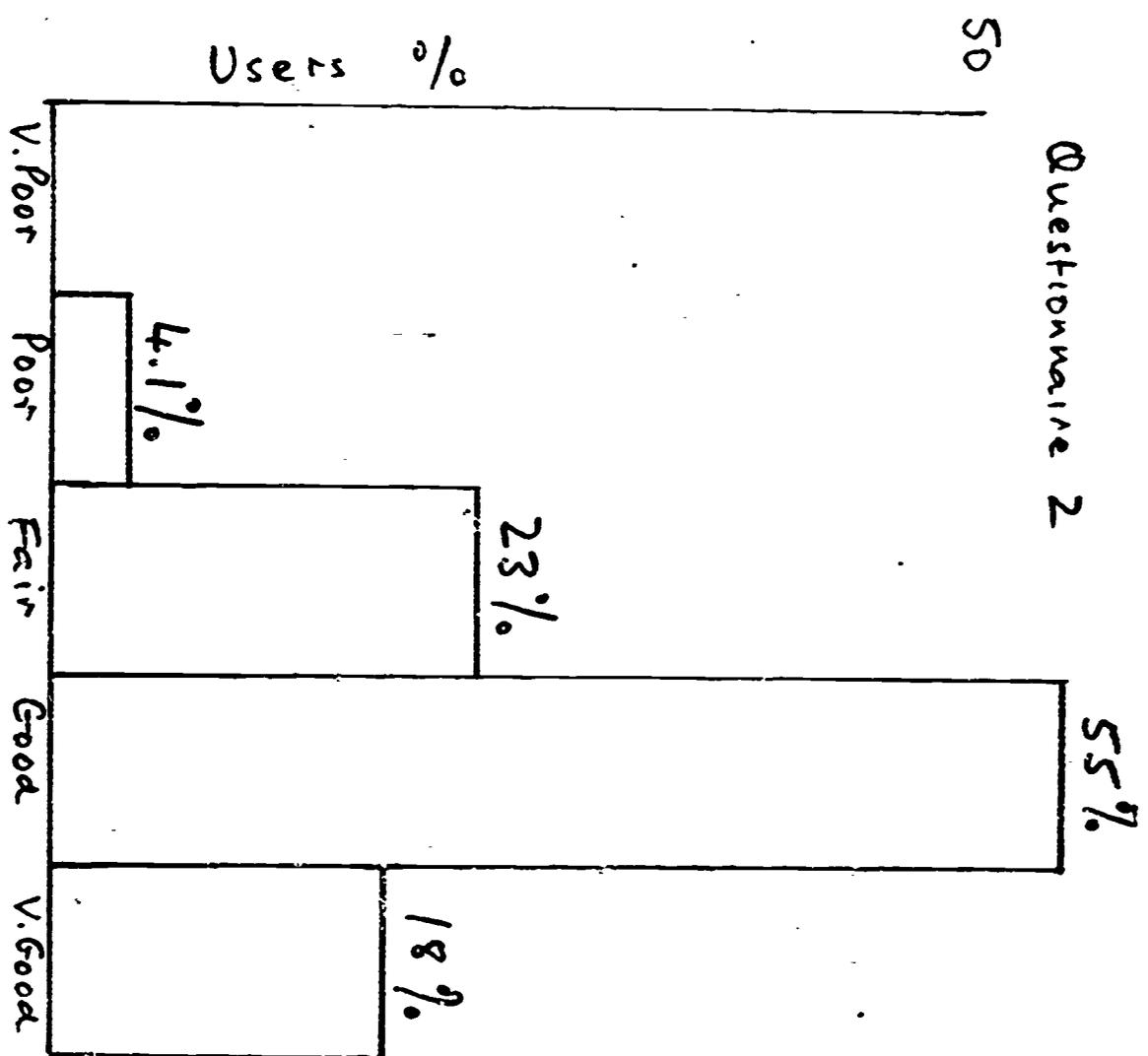
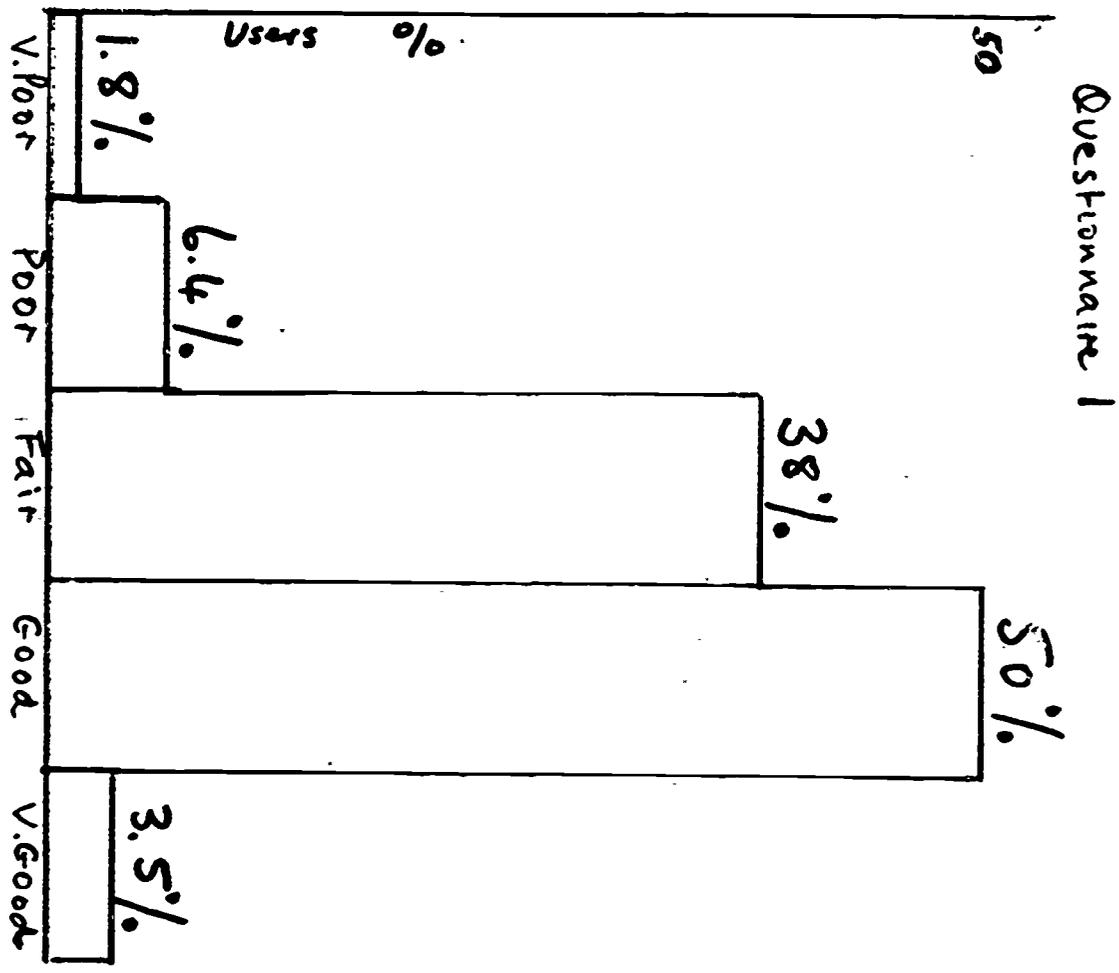


Fig. 3. User satisfaction for Compiler 1

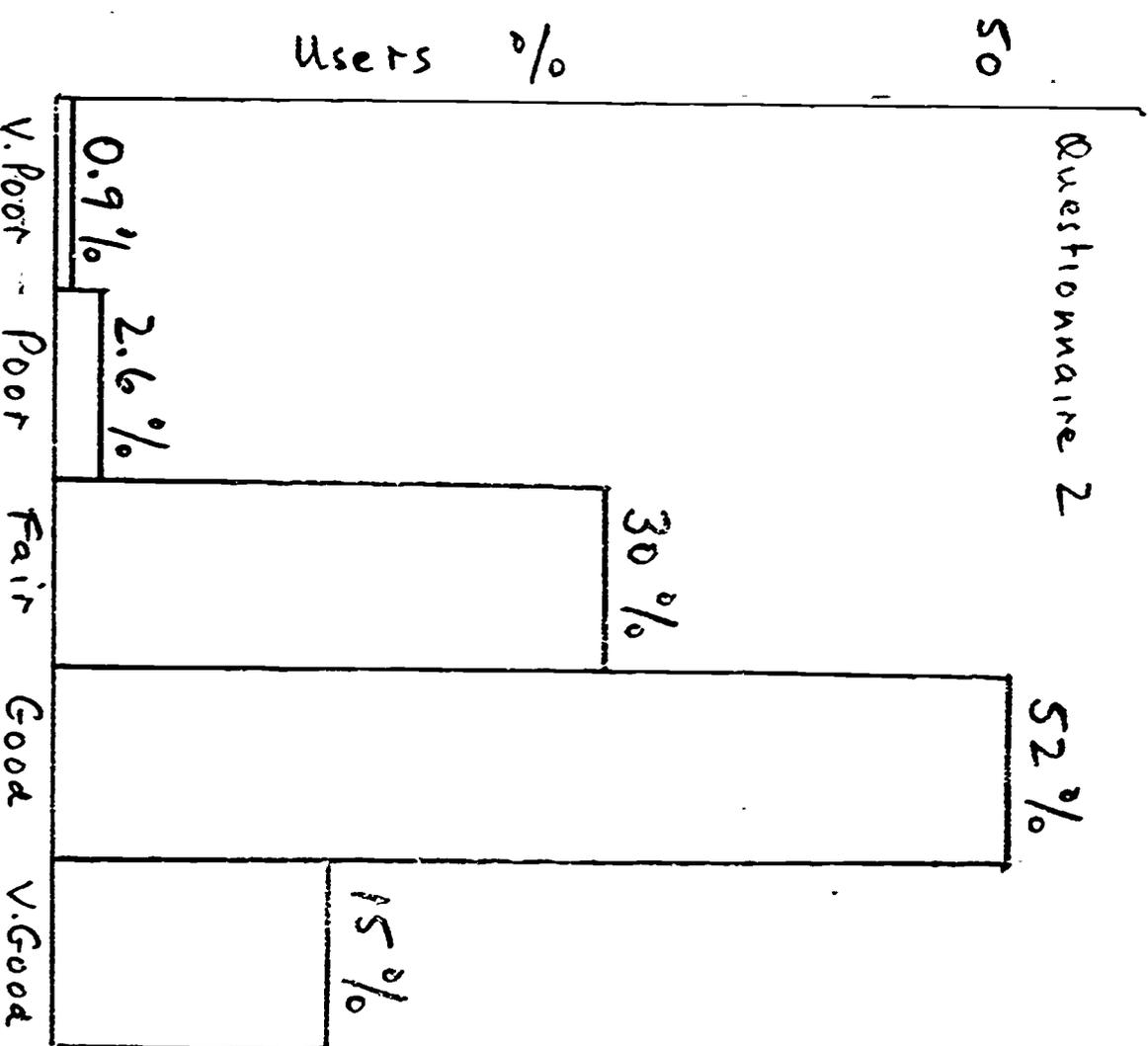
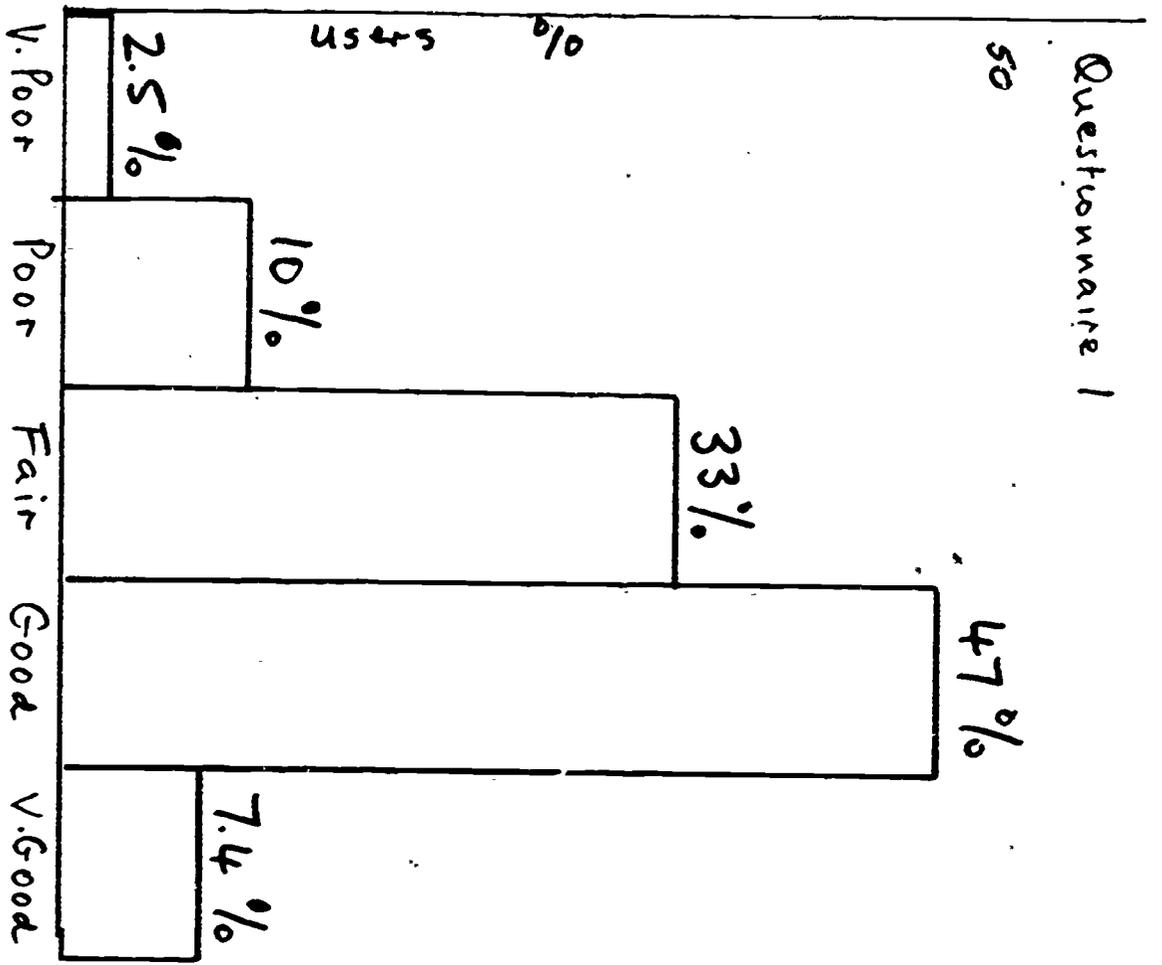


Fig. 4 User satisfaction for Compiler 2

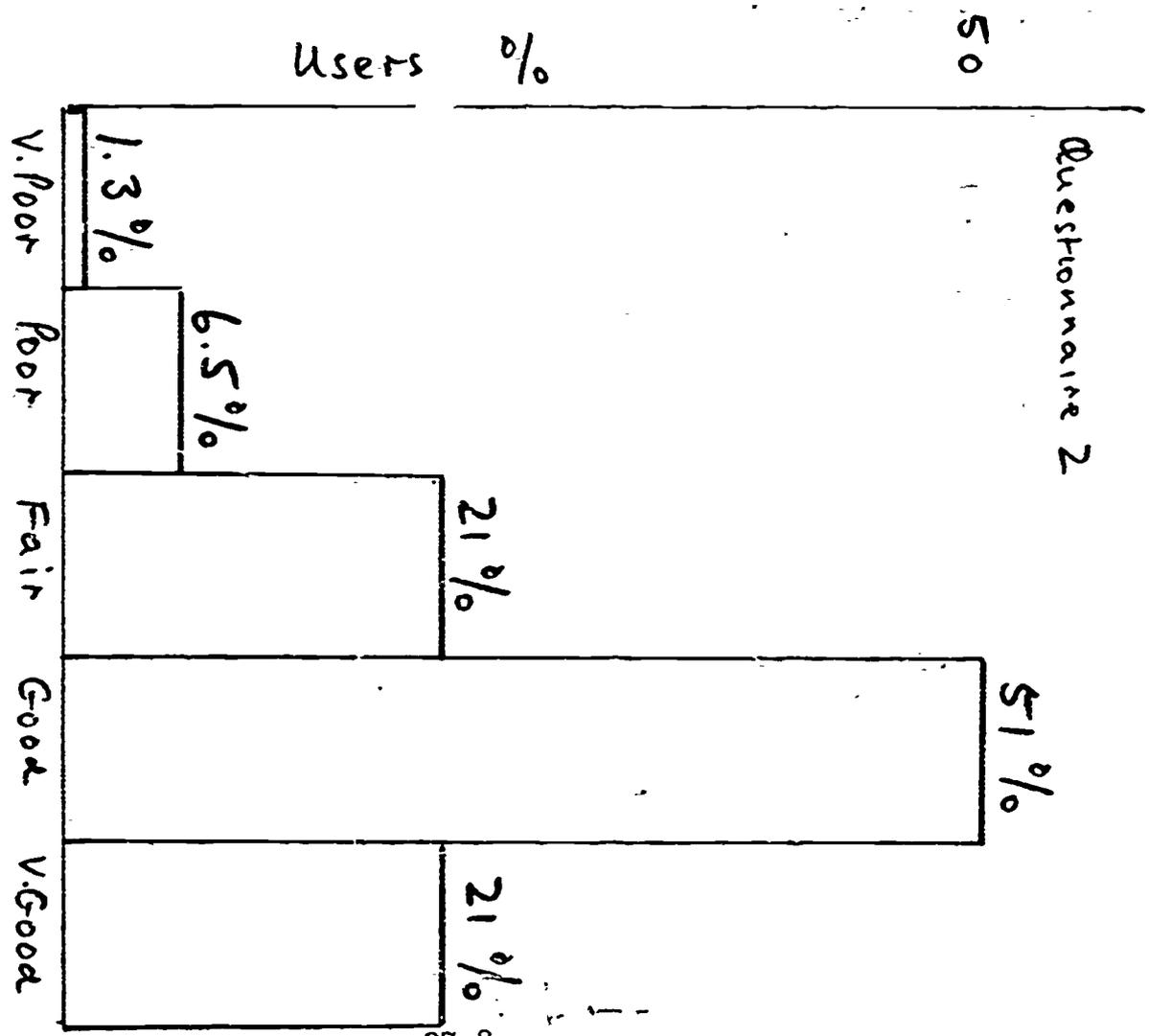
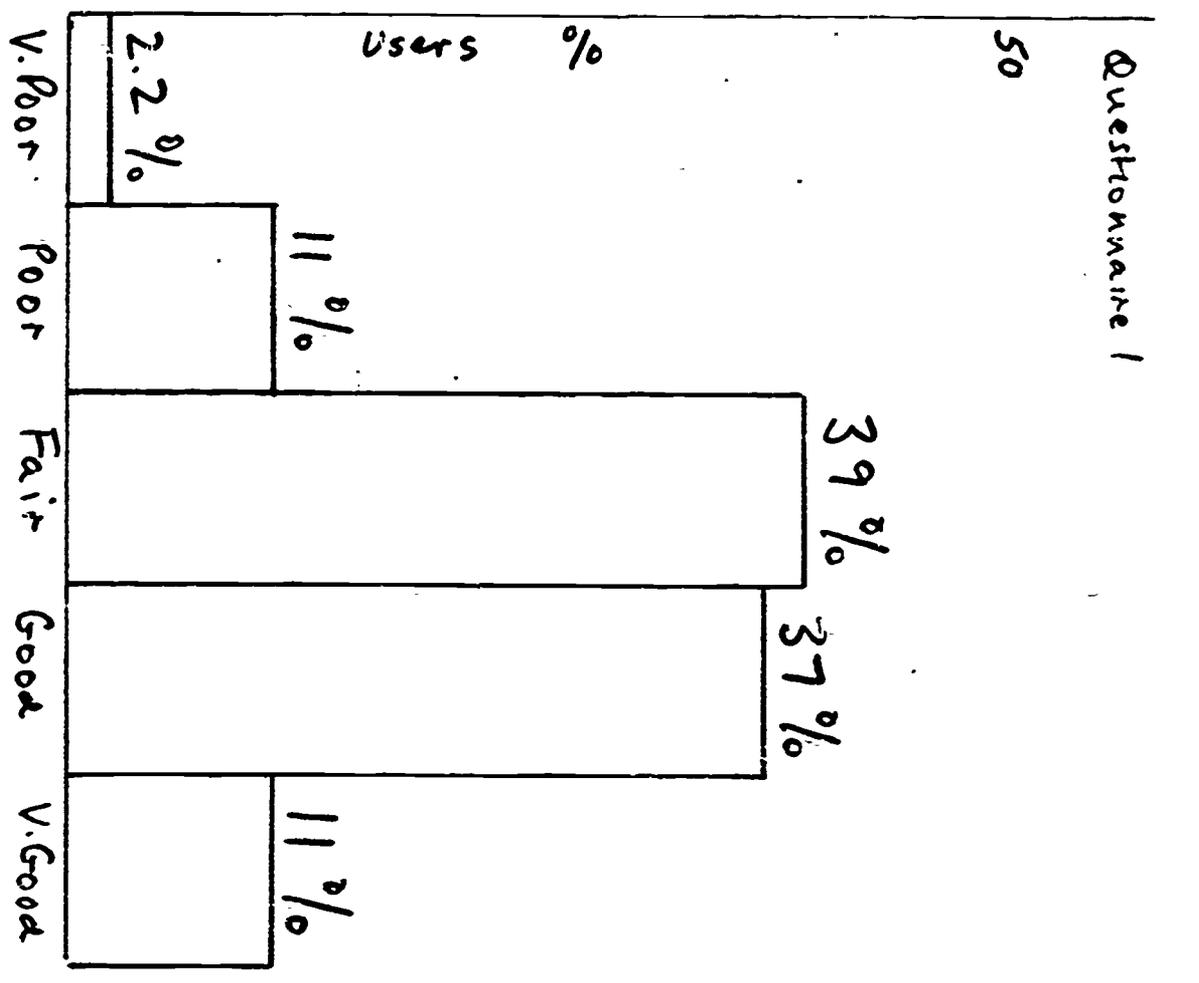


Fig. 5 User satisfaction for Compiler 3

le	Time	Profile	Time								
-		194	30	265	115	332	55	393	125	480	45
60		199	140	267	55	334	60	394	60	481	265
20		200	180	270	90	336	-	398	260	486	65
-		202	120	271	85	337	135	399	70	490	-
-		206	-	276	30	338	190	400	75	491	-
45		207	-	278	5	340	125	401	80	493	-
-		212	50	281	315	343	40	402	105	499	145
155		213	-	284	-	346	-	403	170	500	15
-		216	35	286	160	349	70	405	170	504	80
110		217	300	288	195	350	115	408	140	510	-
35		219	-	289	95	356	200	410	245	514	-
95		222	210	290	60	357	105	419	205	515	-
125		224	45	291	105	359	5	420	95	522	55
175		225	-	298	95	360	20	427	135	524	35
110		231	200	300	-	361	85	430	95	525	-
-		232	180	301	75	362	30	432	-	529	-
15		233	280	303	110	363	15	434	65	531	-
175		237	10	306	-	366	155	435	-	533	75
35		238	60	307	-	367	55	436	70	534	-
30		242	30	309	300	369	105	442	45	544	-
195		243	75	312	95	371	95	444	170	546	170
50		244	85	313	105	379	75	456	125	547	185
180		245	145	316	50	381	70	464	305	552	80
40		246	-	319	25	383	90	466	50	553	75
95		248	15	320	-	384	120	471	40	554	-
140		249	35	323	85	385	165	473	110	555	100
40		256	30	326	65	389	-	477	90	557	245
210		257	130	328	65	391	-	478	-	558	-
100		259	165	329	135	392	35	479	60	566	-

FIG. 6 (contd.)

File	Time
57	20
59	170
70	115
73	30
75	95
77	245
78	80
81	85

Total time = 15385

Av. time = 84 min

183 profiles

Profile	Time								
006	60	114	180	209	-	282	50	355	-
011	15	118	85	210	-	285	35	358	265
014	150	125	155	214	240	287	25	364	50
015	110	126	250	215	160	292	340	365	-
016	75	131	170	220	-	293	5	370	20
021	100	134	195	226	20	295	30	372	55
123	225	137	-	227	75	296	240	373	75
024	20	146	120	228	220	297	90	375	215
031	-	149	90	230	55	302	-	376	60
036	60	151	75	234	35	304	65	377	160
045	-	152	30	236	-	308	-	380	110
046	250	161	10	239	210	310	-	383	75
047	125	162	50	240	45	314	20	386	110
048	105	163	-	241	240	315	-	388	-
057	210	164	155	247	180	321	-	397	240
060	-	165	-	253	235	324	275	404	60
070	105	171	-	254	175	325	40	409	25
073	-	179	375	258	80	330	55	411	302
074	135	182	60	260	100	335	-	416	60
075	150	186	-	261	90	339	-	417	310
082	270	191	165	263	180	341	30	418	-
083	-	193	-	264	30	342	30	422	-
084	125	195	-	266	90	344	135	424	120
086	-	197	150	268	-	348	520	425	185
101	150	201	-	269	120	351	145	426	220
102	60	203	-	272	35	352	260	437	190
104	120	204	195	273	-	353	-	448	130
112	35	208	30	277	75	354	35	450	-



FIG 7 (contd.)

Profile	time	Profile	time
458	40	521	-
459	-	523	55
461	35	526	-
462	40	527	-
468	25	528	30
469	195	530	90
470	-	532	5
475	145	535	120
476	150	536	-
482	15	537	145
485	65	538	95
488	100	539	120
501	110	549	-
502	70	556	-
506	70	559	110
508	-	560	90
509	-	564	160
511	20	565	250
512	120	571	-
513	90	572	270
516	120	576	-
517	190	579	10
518	85	580	135
520	225	582	90
-			

Total time = 17092 min

Av. time = 91 min

188 profiles

Profile	Time								
002	120	062	150	130	40	294	95	540	180
004	270	065	185	132	-	317	60	543	-
005	45	067	300	133	-	318	-	551	30
007	-	072	10	135	-	322	55	562	180
008	25	078	210	138	75	333	120		
012	50	079	55	139	-	347	-		
013	-	088	215	142	90	368	60		
019	-	089	130	145	120	407	490		
026	180	090	-	147	30	412	-		
029	190	092	70	148	330	413	-		
030	-	095	-	153	-	414	315		
032	210	096	40	156	120	415	180		
033	15	097	-	158	-	433	-		
034	-	098	92	159	-	438	220		
039	90	099	45	160	245	440	60		
040	-	100	60	173	60	443	90		
041	60	106	90	176	90	445	105		
042	60	107	-	177	-	446	120		
043	210	109	150	185	375	447	225		
049	60	111	100	187	-	452	180		
052	215	113	40	198	-	453	-		
053	15	117	30	205	455	454	210		
054	90	119	135	218	-	455	-		
055	285	121	90	251	240	460	225		
056	30	122	120	262	210	467	660		
058	150	123	105	279	-	503	190		
059	-	129	270	280	-	505	150		

Total time = 11942 min  
 Av. time = 107 min

## Chapter 28

### METHODS OF USER INTERACTION IN PROFILE COMPILATION

A proposal to include a study of the effect of different types of user interaction was made, in the course of the Investigation, in December 1967. The proposal is given in full in Appendix 28A. The types of interaction compared were:

(A) No interaction

The profile was compiled on the basis of the user's statement of information requirements.

(P) Profile for comments

A draft profile was submitted to the user for comment. Modifications were made, on the basis of his comments, to form the final profile.

(C) Questions

After the user's statement of information requirements had been studied and a draft profile compiled, questions were addressed to the user wherever there was the slightest doubt as to the compiler's understanding of his requirements. The final profile was compiled in the light of his reply.

(D) Interview

The user was interviewed by a member of the SDI staff to discuss his statement of information requirements, its implications, etc. The final profile was compiled on the basis of the statement and interview.

(E) As required

In drafting the profile, the compiler was free to have no interaction, send the profile for comment or submit questions as he considered necessary. Because of the travelling time which would have been involved it was decided that the 'interview' option should not be included.

It was originally intended to continue the comparison of interaction in the profile modification by dividing all the profiles used in the compilation study into two groups, the first of which (X) would have no interaction with the user (other than the normal relevance assessments of the output), and the second (Y) would have any or no interaction as found necessary. However, because of the over-riding need to ensure that as many profiles as possible were satisfactory before the start of the experimental which was expected in April 1968, it was decided to make no differentiation between X and Y profiles and to modify all profiles without interaction with the user.

Selection of groups

To ensure that a number of users were available for interview at a location, the random selection of the five groups, A,B,C,D and E, was made in sequences of five consecutive profile numbers.

A random selection was made, first by type of organisation (ie academic, governmental or industrial), then by organisation, and finally by users within the organisation (in sequences of five users).

At this stage in the investigation a considerable number of profiles had been compiled. This in itself did not necessitate their exclusion from the study since no interaction had taken place and the profiles could be considered draft profiles only, so long as they had not been modified in response to relevance assessments.

However, those profiles which were originally compiled during the first few months of the Investigation were omitted from the study, since this was a learning period during which expertise in profile compilation was built up within the team and, perhaps of more importance for the study, the working thesaurus was under its most intensive development.

Similarly there were a substantial number of profiles which for a variety of reasons were dealt with too late to be included in the study, ie the need to compile them and have them performing satisfactorily at the start of the experimental service made it impossible to carry out any of the interactions with the user. The equivalent "no interaction" profiles were also omitted.

### Implementation of interaction methods

#### Group B - profiles submitted for comment

Since a controlled-language was used, based on a thesaurus to which the user did not have access it was realised that this method of interaction was unlikely to be very productive. The letter explaining what was required of the user (if he wished to take part in the study) is given in Appendix 28B. As may be seen much of it had to be devoted to an explanation of the controlled language and the profile logic.

#### Group C - questions posed by letter

For Group C profiles, the compilers were required to isolate all the points in each statement of information requirements on which there could be any doubt as to the user's meaning or intention. Questions were compiled which sought elucidation on these points and which were sent to the user after review. The covering letter sent to users with the queries is given in Appendix 28C and an example of a question form in Appendix 28D.

#### Group D - interviews

It was considered that the investment in staff time (in travelling etc) was so great for interviews that it was essential that the maximum information should be obtained from each interview. Since, for convenience, all of the Group D users in one location would be interviewed by the same staff member who was unlikely to be the compiler of all these profiles, it was arranged that the compiler would isolate the main queries to be answered in the interview, and that the interviewer would incorporate these queries in preparing the interview structure.

To ensure that a consistent approach was adopted in preparing for the interviews, a skeletal interview plan was produced as shown in Appendix 28E. This was used in the preparatory work to form the basis of the interview questionnaire, an example of which is shown in Appendix 28F.

#### Group E - interaction as required

In general the profiles in Group E were dealt with by one of the above methods, the only difference being in the freedom of the compiler to decide which method of interaction (or none) was most appropriate.

#### Performance Tests

When the profile had been compiled it was tested against the test collection of documents (Test 1). If the performance was considered satisfactory, the profile was left in that form for the experimental service.

Where the performance of the profile was considered inadequate, the relevance assessments were analysed, modifications made to the profile on the basis of the analysis, and the modified profile again matched against the test collection. If the results of this second test (Test 1a) were also considered unsatisfactory the procedure was repeated in Test 1b.

#### Professional effort in compilation

In the compilation and further preparation of profiles for an SDI system there are a number of different professional (as opposed to clerical) activities required. These included, for the system used in the Investigation:-

- (1) assimilation of the information provided by the user in his statement of information requirements
- (2) compilation of draft profile
- (3) interaction with the user
- (4) modification of profile in the light of the interaction

For all of these activities the professional staff effort required for each profile was recorded (as time spent) in accordance with the standard practice of the Investigation. Only for the interaction times were special data required. For Group B profiles the submission of the profile for comment was a clerical operation so that, as with Group A, no professional time was used. For Group C profiles, the elements included in the interaction times comprised the isolation of doubtful points, the framing of suitable questions, and the review of these questions before submission.

## Results

### Performance

In Figure 1 the number of profiles in each group which were submitted to the three tests are shown. As can be seen from that figure and in the alternative display in Figure 4, interviewing is by far the most successful and no interaction the least successful in producing an immediately-satisfactory profile. This is confirmed in results for the performance in Test 1 which is tabulated in Figure 2. to have substantially disappeared (Figure 3) and to be roughly equivalent to the use of questions or the 'as required' choice of interaction methods. The 'no interaction' method continues to have the least satisfactory performance, although the disparity is considerably reduced. A surprising result, however, is the clearly superior performance of the profile interaction method. Since the difference in Test 1 between the profiles which were modified as a result of the participant's comments on his profile and those which were not is minor (and in fact favours the unmodified profiles), the final test performance might reasonably have been expected to equate to that of Group A, ie profiles with no interaction. A possible explanation, in particular for the considerable superiority in Precision, is that the compiler, in considering further modifications after Test 1, felt constrained by the knowledge that the user had signified his acceptance of the profile in its original form and made only the most essential modifications to the profiles in that group, whereas for all other groups the tendency was to make more extensive modifications with a view to improving Recall at the cost of Precision.

### Professional effort

The time spent by professional staff in compilation, interaction and modification before Test 1 is tabulated for each group in Figure 2 and the ranking in Figure 4. As expected the time is least for the non-interaction group (A) and most for the interview group (D) with the profile group (B) the "as required" group (E) and the "questions" group (C) lying between them, in that order.

When the time spent up to the final test is considered (Figure 3) the same ranking is shown but the differences between groups are reduced.

### Discussion and Conclusions

The results of the study show that there is a decided advantage in interaction with the user, since the initial saving in professional effort gained by having no interaction is reduced by the increased effort required subsequently and the level of performance of the profiles does not match that of profiles compiled with user interaction.

It is more difficult to draw general conclusions on the most cost/effective method of interaction because of the special circumstances of the SDI Investigation. The question method of interaction will be generally applicable, whatever the type of SDI system, but the results for the method in which users were asked to comment on their profiles are severely limited in their application by the fact that a controlled language was used and a listing of the vocabulary was not available to the user. Again time spent on the profiles for which the interaction was by interview was especially large because the service was being supplied centrally to users scattered throughout Britain. If the service were being provided in one location or if the profile compilation were being carried out by agents in the same location, the time would be greatly reduced and would be likely to be only slightly greater than that for the method in which the user is questioned by letter.

However the change to a local system of profile compilation would not improve the performance of the profiles or the degree of interaction, whereas the change of an uncontrolled language would increase the opportunity for the user to comment on the profiles.

In general, since the difference in performance between different interaction groups of profiles is small, the most cost effectiveness method would seem to be that in which the professional staff cost is least, ie submission of the profile for comment by the user.

Figure 1

Number of modifications (and subsequent tests) considered necessary to obtain adequate performance

Group (Interaction)	Number of profiles in		
	Test 1	Test 1a (%*)	Test 1b (%*)
A (None)	38	34 (89%)	4
B (Profile)	38	28 (74%)	0
C (Questions)	46	34 (74%)	1
D (Interviews)	33	17 (45%)	1
E (As required)	40	29 (73%)	0

\* Percentage of number of profiles in Test 1.

Figure 2

Time spent by professional staff in profile compilation, interaction and modification before initial test (Test 1) and performance in Test 1

Group (Interaction)	Time (hrs)	Performance	
		Recall (R1)%	Precision (R1/2)%
A (None)	1' 28"	45	48
B total (Profile) modified unmodified	1' 45"	47	56
	1' 46"	46	55
	1' 44"	47	56
C total (Questions) modified unmodified	2' 10"	52	56
	2' 12"	56	53
	2' 03"	34*	66*
D (Interviews)	5' 51"	58	57
E (As required)	1' 54"	49	51

\* Average of only six values.

Figure 3

Time spent by professional staff in profile compilation, interaction and modification to obtain adequate performance, and performance on final experimental test (Test 1, Test 1a or Test 1b)

Group (Interaction)	Time (hrs)	Performance	
		Recall (R1)%	Precision (R1/2)%
A (None)	2' 31"	75	58
B (Profile)	2' 35"	83	67
C (Questions)	3' 08"	78	62
D (Interviews)	6' 38"	79	60
E (As required)	2' 43"	82	59

Figure 4.

Ranking of Groups  
by various measures

<u>Ranking</u>	<u>By</u> <u>percentage</u> <u>given Test</u> <u>I only</u>	<u>Least time</u> <u>spent pre</u> <u>Test I</u>	<u>Perfor-</u> <u>mance in</u> <u>Test I</u>	<u>Least time</u> <u>pre final</u> <u>Test</u>	<u>Perfor-</u> <u>mance</u> <u>in final</u> <u>Test</u>
1 (best)	D (55)	A (1' 28")	D (58/57)	A (2' 31")	B (83/67)
2	E (27)	B (1' 45")	C (52/56)	B (2' 35")	E (82/59)
3	{ B (26)	E (1' 54")	{ E (49/51)	E (2' 43")	{ C (78/62)
4	{ C (26)	C (2' 10")	{ B (47/56)	C (3' 08")	{ D (79/60)
5	A (11)	D (5' 51")	A (45/48)	D (6' 38")	A (75/58)

## Chapter 29

### USE MADE OF SDI NOTIFICATIONS

#### Introduction

The regular feedback of information on the performance of the SDI Service consisted of the weekly relevance returns. These indicated which notifications were relevant in two categories: R1 - highly relevant, and R2 - partially relevant. Non-relevant notifications were marked X. Such returns, of course, give limited information but it was decided to restrict the demands on participants to the basic minimum in order to encourage prompt and regular returns.

To discover a little more of what recipients did with the notifications and in particular how many of the articles notified were read, a special study was made.

#### Method

For one week's notifications each participant was asked to indicate on his relevance return whether he

- a) intended to read the particular article
- or b) would file the card for future use
- or c) would discard the reference

The work was carried out over a period of six weeks, a proportion of the participants in turn being asked for this information.

#### Results

The results are summarized in Table 1

It can be seen from the Table that of the 3086 notifications sent, the recipients intended to read 756 (27%) of them and to file a further 1463 (48%) for future reference. As expected the vast majority of those intended to be read were R1 (highly relevant) documents though a not insignificant number of R2 documents were included.

The majority of the items filed for future reference were, of course, R2 documents though it is interesting to find that some 155 items marked as non-relevant were in fact filed for future reference. Without more evidence it is impossible to say whether these represented useful information not directly relevant to the profile or whether cards were being filed indiscriminately. In this connection it is encouraging to find that most of the non-relevant items were being discarded.

## Chapter 30

### Desirability of including material other than English-language periodical articles in the coverage of the SDI Service

The SDI Service provided in the Investigation was restricted in coverage to English-language periodical articles.

In the Final Questionnaire of the Investigation participants were asked to indicate how desirable they thought it that the SDI Service should cover other sources of information. The possible sources were listed as follows:-

1. Books
2. Conference proceedings and papers
3. Foreign-language periodical articles
4. Manufacturers' literature
5. Patents
6. Reports
7. Standards and specifications
8. Theses and dissertations

To each class of material participants were asked to assign the values 0, 1 or 2, having the following meanings:-

- 0 - Don't mind if omitted
- 1 - Inclusion desirable
- 2 - Inclusion essential

### Results

The results are tabulated below. It will be remembered that the participants in the Investigation were, in the main, randomly-selected individuals drawn roughly equally from universities and colleges, government establishments and industrial firms. The breakdown for these three types of user is shown in the table. The 'Miscellaneous' group consists of Group users and some non-randomly selected individuals.

The weighted totals are arrived at simply by counting 0 for the category 0 items, 1 for the category 1 items, and 2 for the category 2 items. From these figures the sources may be ranked in the order-of preference for coverage.

<u>Source</u>	<u>Score</u>	<u>Rank</u>
Conferences	422	1
Foreign periodicals	315	2
Reports	295	3
Theses	217	4
Books	176	5
Patents	162	6
Manufacturers' literature	121	7
Standards and Specifications	101	8

The same rankings occur if only the figures for the category 2 items are taken, ie those which were considered essential for inclusion. This is convenient since it avoids any discussion of the relative weight to be assigned to category 1 and category 2 items.

It is, of course, also possible to consider the figures from the opposite point of view, ie which items were considered most expendable by the participants. If we do this we find, as might be expected, the reverse order of ranking from that given in figure 2, with one exception: foreign periodicals are considered more expendable than reports.

If we compare the rankings assigned by university, government and industrial participants there are one or two differences but these consist mainly in university participants' placing, as might have been expected, a high value on thesis material, and industry rating patent literature rather more highly.

#### Rank

	<u>University</u>	<u>Government</u>	<u>Industry</u>
Conferences	1	1	1
Foreign periodical	2	2	3
Reports	4	3	2
Theses	3	4	7
Books	5	5	5
Patents	6	7	4
Manufacturers' literature	8	6	6
Standards and Specifications	7	8	8

Table 1

	Books	Conferences	Foreign Periodicals	Manufacturers' Literature	Patents	Reports	Standards	Thesis	Total
<u>Universities</u>									
2	15	41	25	2	7	16	2	20	128
1	28	41	52	14	29	46	19	53	282
0	48	10	17	76	56	28	71	19	325
Weighted total	58	123	102	18	43	78	23	93	
<u>Government Establishments</u>									
2	8	44	29	10	3	20	6	11	136
1	34	43	41	28	24	57	22	33	282
0	62	19	36	66	72	27	75	60	417
Weighted total	50	131	99	48	40	97	34	55	
<u>Industry</u>									
2	7	36	14	9	6	18	3	3	96
1	32	49	50	26	42	54	26	35	314
0	63	17	38	66	55	26	73	64	402
Weighted total	46	121	78	44	54	90	32	41	
<u>Miscellaneous</u>									
2	7	17	8	2	6	8	1	7	56
1		13	20	7	13	14	10	14	103
0	15	4	7	25	15	11	22	12	111
Weighted total	26	47	36	11	25	30	12	28	
<u>Total</u>									
2	37	138	76	23	27	62	12	41	416
1	102	146	163	75	108	171	77	135	977
0	188	50	98	233	198	92	241	155	1255
Weighted total	176	422	315	121	162	295	101	217	

## Chapter 31

### NOVELTY OF INFORMATION PROVIDED BY THE SDI SERVICE

#### Introduction

For any subscriber to an SDI or any other current-awareness service, the notifications he receives from that service will form only one source of information on relevant articles. Many subscribers will regularly see the two, three or more journals central to their field of interest and which regularly contain relevant articles. Most of the articles in these journals will, therefore, already be known to the subscriber before the SDI service can hope to notify him of them. Inclusion of such articles among his notifications will serve possibly to confirm that the service is effective, but he is unlikely to be willing to pay for a service which only serves to tell him what he already knows. An important factor in the performance of any SDI service is the "novelty" of the information provided, i.e. the extent to which notifications draw the subscriber's attention to information which he does not otherwise see.

An attempt was made to assess this in the SDI Investigation.

#### Method

For one week's notifications each participant was asked to give, in addition to his normal relevance assessment for each article notified, an indication of whether he had been aware of its existence previously. To spread the load on the SDI staff this study was done over a period of six weeks.

#### Results

The results are shown in Table 1

Table 1 Novelty of information provided by the SDI Service

2	3	4	5	6	7	8	9	10	11	12	
No. of Users	Number of Notifications sent	Relevant sent			Relevant not previously seen						
		R1	R2	R1+R2	R1		R2		R1+R2		
		No.	%	No.	%	No.	%	No.	%		
1	90	623	237	217	454	156	66	183	84	339	75
5	87	476	194	311	505	144	74	268	86	412	81
5	26	145	42	40	82	30	71	33	83	63	77
7	58	481	163	186	349	135	83	166	89	301	86
3	71	608	178	234	412	137	77	205	88	342	83
9	<u>30</u>	<u>235</u>	<u>66</u>	<u>105</u>	<u>171</u>	<u>52</u>	<u>79</u>	<u>90</u>	<u>89</u>	<u>142</u>	<u>83</u>
a1	362	2828	880	1093	1973	654	74	945	86	1599	81

Columns 1 - 3 show for each week the number of participants involved and the total number of notifications they received. Columns 4 - 6 show the number of these notifications which were highly relevant (R1) and of secondary relevance (R2). Columns 7, 9 and 11 give the numbers of R1 and R2 notifications referring to articles of which the recipient was not previously aware. Columns 8, 10 and 12 show these as a percentage of the relevant notifications sent in each category, R1, R2 and R1+R2.

It can be seen from the table that 74 percent of the high relevance (R1) articles notified by the SDI Service had not previously been seen by participants. The figure for articles marked R2 (partially relevant) is somewhat higher i.e. 86 percent, giving an overall figure for both R1 and R2 articles of 81 percent.

"Useful" Precision

On the basis of the figures in the table it is, of course, possible to arrive at figures for the "Useful" precision performance of the system, i.e. relevant notifications not previously seen expressed as a percentage of total notifications.

<u>Notifications</u>	<u>Relevant</u>		<u>Not Previously</u> <u>seen</u>		<u>Precision</u>		<u>Useful</u> <u>Precision</u>	
	<u>R1</u>	<u>R1+R2</u>	<u>R1</u>	<u>R1+R2</u>	<u>R1</u>	<u>R1+R2</u>	<u>R1</u>	<u>R1+R2</u>
2828	880	1973	654	1599	31%	70%	23%	57%

## Chapter 32

### ACCEPTABILITY TO USERS OF ALTERNATIVE FORM OF SDI NOTIFICATIONS

#### Introduction

In designing the SDI Service it was decided that the notifications should be sent in the form of cards. Of the possible standard sizes, 5" x 3", 6" x 4", and 8" x 5", it was decided to adopt 6" x 4" since the small card placed too tight a restriction on the amount of information that could be included. The 8" x 5" card was considered rather larger than required and such information as we had indicated that potential participants in the Investigation used this size less than the smaller cards.

However, although cards were thought likely to be the most flexible and convenient form of notification a study was planned as part of the Investigation to discover whether an alternative form of notification would be acceptable to recipients.

The obvious alternative to the cards was a computer-produced listing of matching documents giving the author, title, citation and descriptors.

#### Method

It was originally planned to provide the alternative form of notification for four consecutive weeks during the middle of the twelve-months operational period. However, owing to lengthy delays in obtaining minor adjustments to the layout of the matching output, it was not possible to provide more than one week's notifications in this form before the end of this period of the Investigation in December 1969. The lateness of this study had perhaps one advantage in that it occurred after recipients had been informed of the imminent changeover to a cost-recovery service. Thus there was an incentive for people to consider seriously the question of preferred form of notifications since they would shortly be paying for the service.

For the week in question (Week number - series 063) recipients were sent a printout showing details of the documents matching their profiles instead of the set of cards normally sent. The accompanying questionnaire (appendix 32A) asked users to indicate their views on the new form of notifications as follows:

- a) Very much prefer the new form of notification
- b) Somewhat prefer the new form of notification
- c) Very much prefer the cards
- d) Somewhat prefer the cards
- e) Have no particular preference

The questionnaire also asked the reasons for any particular preference.

#### Reasons for preferred form of notifications

The SDI participants, in addition to stating their preference, were asked to give their reasons briefly.

The reasons given covered various merits and defects of the two types of notification, e.g. readability, ease of handling, etc, and though these were expressed in different ways it was possible to group them.

The reasons for preferring cards are summarized in Table 3. The groupings obviously overlap since the flexibility associated with one record per card is closely connected with the argument for ease of filing. However, the headings service to group the reasons stated or clearly implied by the users.

Results

Of the 437 participants who were sent notifications in this form 359 sent in replies. The results are shown in Table 1.

Table 1. User preference for cards vs. paper print-out

<u>CARDS</u>		<u>LISTING</u>		<u>No Preference</u>	<u>Total</u>
<u>Very much prefer</u>	<u>Somewhat prefer</u>	<u>Very much prefer</u>	<u>Somewhat prefer</u>		
257 (72%)	63 (18%)	5 (1.4%)	17 (4.7%)	17 (4.7%)	359

A preference for cards was expected, but the almost unanimous preference did cause some surprise. A breakdown by type of organisation is shown in Table 2.

Table 2 Preference for cards by type of organisation

<u>Type of User</u>	<u>Very much prefer</u>	<u>Somewhat prefer</u>	<u>Very much prefer</u>	<u>Somewhat prefer</u>	<u>No Preference</u>	<u>Total</u>
University	78	18	2	-	5	103
Government	81	23	1	8	5	118
Industry	78	20	1	7	6	112
Group	20	2	1	2	1	26
<b>Total</b>	<u>257</u>	<u>63</u>	<u>5</u>	<u>17</u>	<u>17</u>	<u>359</u>

Table 3

<u>Reasons in favour of cards</u>	<u>Number of people mentioning this factor</u>
<u>FILING</u>	
More convenient for filing and later reference	253
Compatible with normal filing system	23
Assists further sub-classification by the user	29
<u>HANDLING</u>	
Less liable to damage or wear	8
Easier to handle and read	20
<u>FLEXIBILITY</u>	
Easier to pass individual references to colleagues	34
Convenient to have one card for each individual paper	14
Easier for taking to library to look up relevant articles	10
Convenient for use in requesting loan	9
<u>RECORDING OF ADDITIONAL INFORMATION</u>	
Summary of the article can be written on the back	10
There is space for more detailed abstract	4
Additional information can be added	10
<u>OTHER REASONS</u>	
Have become used to cards, inconvenient to change	2
Family like to write on the dud cards at home	1

Table 3 (continued)

<u>Reasons in favour of cards .</u>	<u>Number of people mentioning This factor</u>
<u>LEGIBILITY</u>  Cards easier to read, better layout of information, upper and lower case easier to read than line-printer output.	40
<u>DISCARDING IRRELEVANT ITEMS</u>  Convenient to discard irrelevant material or give special attention to urgent items.	55

Table 4

<u>Reasons for preferring Computer Listing</u>	<u>Number of People Mentioning This factor</u>
<u>FILING</u>	
Easier to file	7
Occupy less file space since fewer individual pieces of paper	3
Adaptable to folder storage	1
<u>HANDLING</u>	
Easier to handle	1
<u>FLEXIBILITY</u>	
Easier to take to library	1
<u>LEGIBILITY</u>	
Easier to scan folder of pages than individual cards	3
Easier to read and understand	3
Easier to see at a glance	2
<u>RELEVANCE RETURNS</u>	
Easier for assessment of relevance	5
Self-copying form	2

Table 5

<u>Disadvantage of either form of notification</u>	<u>Number of People mentioning this factor</u>
<u>CARDS</u>	
More easily lost	1
Wrong size, should be 5" x 7"	1
<u>PRINTOUT</u>	
Difficult to file	5
Not so easily stored	3
Sensible filing system impossible	1
Filing by subject not possible	6
Would involve transfer of information to cards for filing	7
Very inconvenient to keep	4
Great deal of unused paper	1
Inconvenient size, incompatible with other paper formats	1
More difficult to read	5
Difficult to scan	5
More easily torn	1
Irrelevant items cannot be discarded	1
Individual items cannot be passed to members of research group - causing delay	2

## CONCLUSIONS

The results appear overwhelmingly to justify the original choice of cards as the form of notifications. Very few people (6.3%) showed any preference for the listing and only 1.4% had a strong preference. Strong preference for the cards on the other hand was shown by 71% of the users. The reasons for this preference appear generally agreed. These are that the cards provide a conveniently-handled, individual record for each item which can be filed, passed to other interested people or discarded as required.

Some other reasons for preferring the cards must be treated in context, particularly that of greater legibility. Partly this is due to the fact that the cards use upper and lower case type as against the upper case only of the computer printout. On the other hand the layout of the information on the printout could have been improved, in particular by separating the descriptors from the journal reference.

It must also be said that the printout was a new form imposed for one week only on an established system. As some people indicated, they had adapted to the cards and the printout brought problems of compatibility.

### Miscellaneous comments

A number of the replies contained comments of a general nature of a dealing with points somewhat outside the immediate purpose of the study. Since they include useful suggestions for improving the service they are given below.

"One or two of the recent cards contained abstracts which tended to be more informative than just keywords.

The keywords section might be clearer for sorting if dropped one line from the title.

I shall withdraw from the scheme if the new form of notification is introduced.

It is possible to include brief abstracts.

Suggest print authors first: I think majority of people file by authors.

Within a few years most companies will have a time-shared graphic computer terminal. The new form of presentation would then be valid if the references could later be accessed from a central computer store.

A format within 5" x 2" printed on thin paper for sticking to cards in a filing system would be most welcome.

Punched classified cards would be ideal.

There is a general problem that many titles of articles, particularly American, are not an immediate indication of contents. A very brief synopsis in such cases would give some meaning.

"Title of Paper" information is often insufficient to determine relevance of technical content. The difficulties or providing further information is appreciated but one wonders what effect this may have had on the results of this study.

There would be some advantage in attempting to give references the classification used in Current Papers and Abstracts.

I would be inclined not to use the service unless cards were at least available on request.

If the references were available on IBM punched cards the value of the service would be increased tremendously.

A sensible filing system for the references, eg by topic is impossible with this new form of notification. The SDI service has been useful in providing 1) an up-to-date appraisal of the literature, 2) a permanent record of the literature. Adoption of this new form of notifications would seriously impair the usefulness of the service to me.

A useful addition would be the place of origin, since the authors are not always known, and journals not always immediately available elsewhere.

Sometime ago you were able to append short abstracts to the references. This increased the value of the service enormously. I would still feel it worthwhile if only, say, 30% - 40% of references had abstracts even if the cost of the service (to the user) went up by about 10% - 20%.

The language in which the article is available would be of interest if noted, also if available in translation, for instance "in Russian, English edition available".

The address of the authors, so that reprints of important articles can be written for without having to consult the actual publication, should be printed on the cards.

The card system of presentation is one of the main benefits of the SDI Scheme as far as I am concerned.

It would be helpful to have the date of printing or delivery on the cards."