Tests of indexes. A comparison of keyword from title indexes with and without added keywords and a single access point per document alphabetic subject index*

Report by:
G. Jahoda
Mary Lou Stursa

Work by:
Mary Culnan
G. Jahoda
John Lattimer
Mary Lou Stursa

January 1969

School of Library Science
Florida State University
Tallahassee, Florida 32306

*Work supported in part by the Air Force Office of Scientific Research, Office of Aerospace Research, United States Air Force, under AFOSR grant number 895-65.

.61

Table of Contents

List of Figures

List of Tables

Abstract

Three indexes to a collection of 3204 documents in the field of chemistry were test searched. The indexes are a keyword from title index without added keywords, a keyword from title index with added keywords, and a single access point per document alphabetic subject index. The three indexes were searched by 14 graduate chemistry students with 66 questions. Search results are characterized in terms of recall, precision, search time, and three other single figure measures. A measure of index performance based on recall, precision and search time was developed. There is no statistically significant difference between search results with the multiple access points per document keyword from title index and the single access point per document alphabetic subject index. Statistically significant differences in search results were found between the keyword from title index with added keywords and the keyword from title index without added keywords. The effect of the size of the delete word list and of elements of vocabulary control in keyword from title indexes were also studied.

I. Introduction

Several hundred records of researchers' document use have been collected as part of a study of information gathering habits. Thirteen researchers in science and engineering at Florida State University participated in this study of use of documents in the researchers' personal document collections. Each record of use consists of a record of the question and a record of the bibliographic citation of documents that the researcher considered relevant for the question. The case histories of

document use were collected to gain additional insight into information gathering habits of researchers and also to determine whether information specialists can be of assistance in the preparation of indexes to researchers' personal document collections. For further background on this aspect of the study, the reader is referred to a publication by Jahoda, Hutchins, and Miller.[1] Six keyword from title indexes have been prepared for the personal document collections of six researchers. Three of these indexes, the indexes to the document collections of three chemists, are used in the tests of indexes that are described in this report. These tests were intended to answer the following questions for indexes to about 3200 documents:

* What is the effect of index depth on index performance?

* What is the effect of added keywords, size of delete list, and elements of vocabulary control on the performance of keyword from title indexes?

II. Description of work

a. The size of the document collection and the index variables tested

The derivation of the five indexes used in the test and their characteristics are outlined in Fig. 1. Each of the five indexes is to the same collection of 3204 documents. These 3204 documents represent the merged document collections from the offices of three chemists for whom keyword out of context (KWOC) indexes have been prepared. Two of the three chemists added keywords to titles in their KWOC indexes; the third chemist did not. Each of the chemists selected his own list of delete words. The three individual lists were merged into one list of 628 delete words which was used in the preparation of the merged KWOC index. The

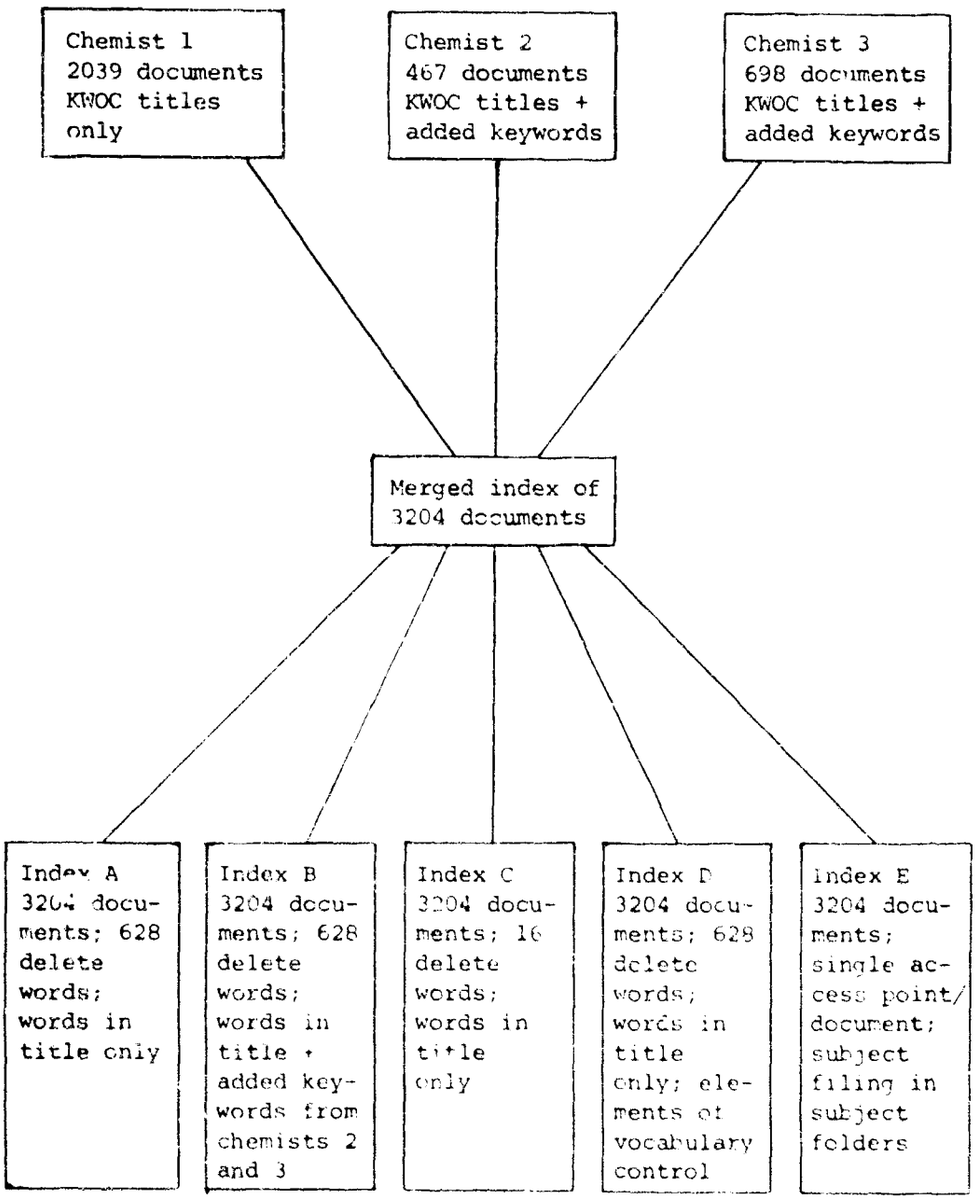


Fig. 1. Characterization of indexes used in tests

merging of the three chemists' KWOC indexes (two of which had additional keywords) yielded Index B. Index A, the index that is used as the basis of comparison against the other four indexes, was prepared by removing the additional keywords from Index B. Index A is thus the same as Index B except that it does not have the added keywords.

Index C was to have been the same as Index A but it was to have a minimum delete word list of 16 words instead of the list of 628 delete words in Indexes A, B and D. The list of delete words in Index C are the following 16 words which have been used by H. P. Luhn in the preparation of one of his keyword from title indexes:

a, an, and, as,at, by, for, from, if, in, of, on, or, the, to, with.[2]

The size of the delete word list is therefore the variable to be tested in Index C.

Index D differs from Index A in that it was to contain the following elements of vocabulary control, the variable to be tested in this index:

* Lumping of singular and plural forms of nouns
* Lumping of words that differ from each other only in word form, e. g., analysis, analyzing, analyzed
* "See" cross-references for synonyms
* "See also" cross-references for word relationships other than specific-generic relationships, e. g., ion-exchange and cation exchange.

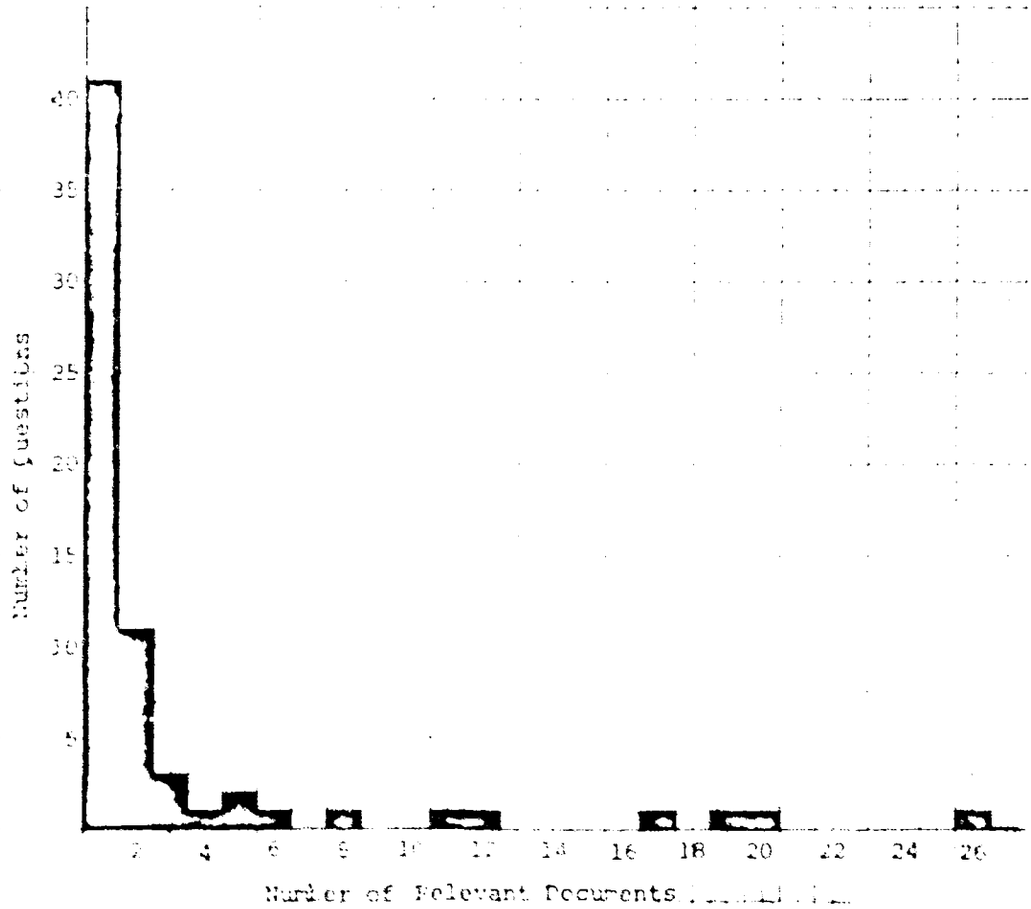
Each of the 3204 documents included in the KWOC index is filed in a subject folder in one of the three chemists' offices. The list of these subject folder headings may be considered as the index headings of an alphabetic subject index. Index E is a single access point per document

alphabetic subject index. It consists of an alphabet listing of the subject headings under which the documents are filed along with the bibliographic citation (including title) of each document so filed.

b. The questions searched and the search procedure

The indexes were searched with 66 questions. The questions were selected from the case histories of use of the documents in the three chemists' offices. Not all of the three chemists' recorded case histories were used in the test. Questions that were for the recall of a specific document, questions that were almost identical to questions used, and questions that were primarily for general background rather than for everything in the files on a subject were all omitted from the test. The case histories (the question and the document or documents considered relevant by the chemist searching his own collection) were collected before the individual KWOC indexes had been prepared. The list of relevant documents for each question was reviewed by the chemists before the tests were started. During this review the chemists selected additional relevant documents that were either not in the collection or were missed when the original searches were made. The 66 questions used in the test are listed in Appendix A. Fig. 2 gives a tabulation of the number of relevant documents per question.

Three of the indexes, Indexes A, B and E, were searched by 14 graduate chemistry students. The searchers were given both written and oral instructions on the use of the indexes before they started the test searches. The written instructions are reproduced as Appendix B. Each searcher also conducted a trial search under the supervision of a member of the project staff. Search results in the trial search were not counted in


Number of Questions
Number of Relevant Documents
4
6
8
16
18
20
22
24
26

the overall test results. The searchers were instructed to search each of the 66 questions in either Index A, Index B, or Index E and to attempt to locate all of the relevant documents for each question. One of the searchers did not complete all of the questions. One other searcher was substantially better than the other searchers in his group as can be seen from Table 1. This searcher's results were therefore excluded from the overall tabulation.

Table 1

Comparison of search results of Group I searchers and excluded searcher

	Group I mean	Excluded searcher mean
Recall	46.0%	67.8%
Precision	26.4%	56.0%
Search time	15.22 min.	9.91 min.

All of the 66 questions were searched in Index A, the control index. Eleven out of 66 questions were also searched in Index B. These 11 questions represented 11 out of 48 questions that could have been searched in Index B since 48 out of the 66 questions were searched by the two chemists who added keywords to their document titles. These 11 questions were selected because they had relevant documents that could be retrieved only or more easily with the keywords added by the chemists. The remaining 55 questions were also searched in Index E.

The searchers were divided into two groups, Group I and Group II. A different search pattern was assigned to each of the two groups. Half of the searchers, the Group I searchers, searched half (33) of the questions in Index A. The other half of the searchers, Group II, searched the other half of the questions in Index A. Group I searched the questions that it did not search in Index A in either Index B or Index E. Group II searched the 33 questions that it did not search in Index A in Index B or Index E. Thus the same questions that were searched by Group I in Index A were searched by Group II in Index B or Index E and vice versa. Each group searched half of the Index B questions in Index B and half of the Index E questions in Index E. Each searcher therefore searched 33 questions in Index A, five or six questions in Index B and 27 or 28 questions in Index E. The search pattern for Group I and Group II is given in Table 2. This search pattern was intended to yield seven complete sets of questions searched in Index A and seven complete questions searched in Indexes B and E. Less than seven complete sets were actually obtained since not all anticipated search results were obtained, for reasons that have already been indicated.

c. Measures of search results

The commonly used recall (number of relevant documents retrieved over the number of relevant documents in collection) and precision (number of relevant documents retrieved over total number of documents retrieved) index measures were used in the test. Relevance judgment by the original searchers, the chemists who conducted the searches in their own document collections and for their own use, were used in the calculation of recall and precision. Each of the searchers in the test was also

asked to keep a record of search time for each question.

Table 2

Search Pattern for Group I and Group II

	Indexes searched			
	A	B	A	E
Group I searchers	questions 1-6	questions 7-11	questions 12-17, 24-42	questions 18-23, 43-66
Group II searchers	questions 7-11	questions 1-6	questions 18-23, 43-66	questions 12-17, 24-42

The commonly used recall and precision measures are less than satisfactory for the comparison of indexes. For example, Fairthorne points out that neither the recall nor the precision ratio tell how many relevant documents were retrieved.[3] Yet a 20% precision may be perfectly acceptable when two relevant documents are retrieved and unacceptable when 20 relevant documents are retrieved. The comparison of indexes by means of curves of recall versus precision is not as useful as might be desired. Is an index with 50% recall and 80% precision better than an index with 40% recall and 90% precision? We do not really know. Search results were therefore also characterized in terms of three single figure measurements to determine whether or not they are more satisfactory than

the recall and precision measures.

The first single figure measure used is called E (efficiency) by Swets, who bases his measure on statistical-decision theory. The conditional probability of a hit is plotted against the conditional probability of a false drop and measured by means of an operating characteristic, the normal deviate to obtain E.[4]

The second single figure measure is called ERF (expected search length reduction factor). The measure, which was developed by Cooper, is based on the calculation of the expected number of non-relevant documents which would have to be looked at before the relevant documents could be found. Retrieval performance of an index is evaluated against random searching through a document file.[5] For both E and ERF measures, the higher the measure the better the index.

The actual time required for conducting searches in an index appears to be an important index variable and one that has not as yet been studied to any extent. If, for example, one index is better in terms of precision but worse in terms of search time than another index, the advantage of better precision may be cancelled by the disadvantage of slower search time. A single figure index measure and one that is based on recall, precision as well as actual search time was therefore sought. In the course of looking for such an index measure, a literature search was made and a potentially relevant document was located. The document is by Lazorick and Minder and it is on the least cost searching sequence for verifying orders of books in a library[6]. The person verifying orders has the choice of several tools or

bibliographies to do this task. He decides on one tool first, and if this tool does not have the bibliographic citation of the requested item, he must go to a second, third, and subsequent tool until he finds the wanted information. Lazorick and Minder suggest that the choice of the first tool and of the subsequent tools (the search sequence) should be based on the time that it takes to search a tool as well as on the likelihood of locating the sought bibliographic citation in that tool. This is expressed as the following ratio:

$$\text{Searching sequence ratio} = \frac{\text{Time for searching a tool}}{\text{Probability of success}}$$

The tool with the lowest searching sequence ratio is the one to search first, the tool with the next lowest ratio is to be searched second, etc.

There seem to be some similarities between the best bibliography for locating and verifying a bibliographic citation and the best index for locating a set of documents. In both cases we are interested in a high probability of success (high recall) to be obtained in as short a search time as possible. There are also differences between these two types of searches, the most important one being that in the case of the bibliography one and only one document is sought and it is either found or not found. In index searches, one or more documents are sought and in addition to retrieving the wanted documents, non-wanted documents (false drops) may be retrieved also. The search time and the probability of success (recall) in the searching sequence ratio can therefore be used for characterizing part of the search results but an additional factor must be added for the retrieval of non-wanted documents, the false drops. We can do this by assigning a demerit for each false drop if we argue

that false drops add to the searching time of an index. The modified searching sequence ratio, called time-recall factor (TRF) is given below:

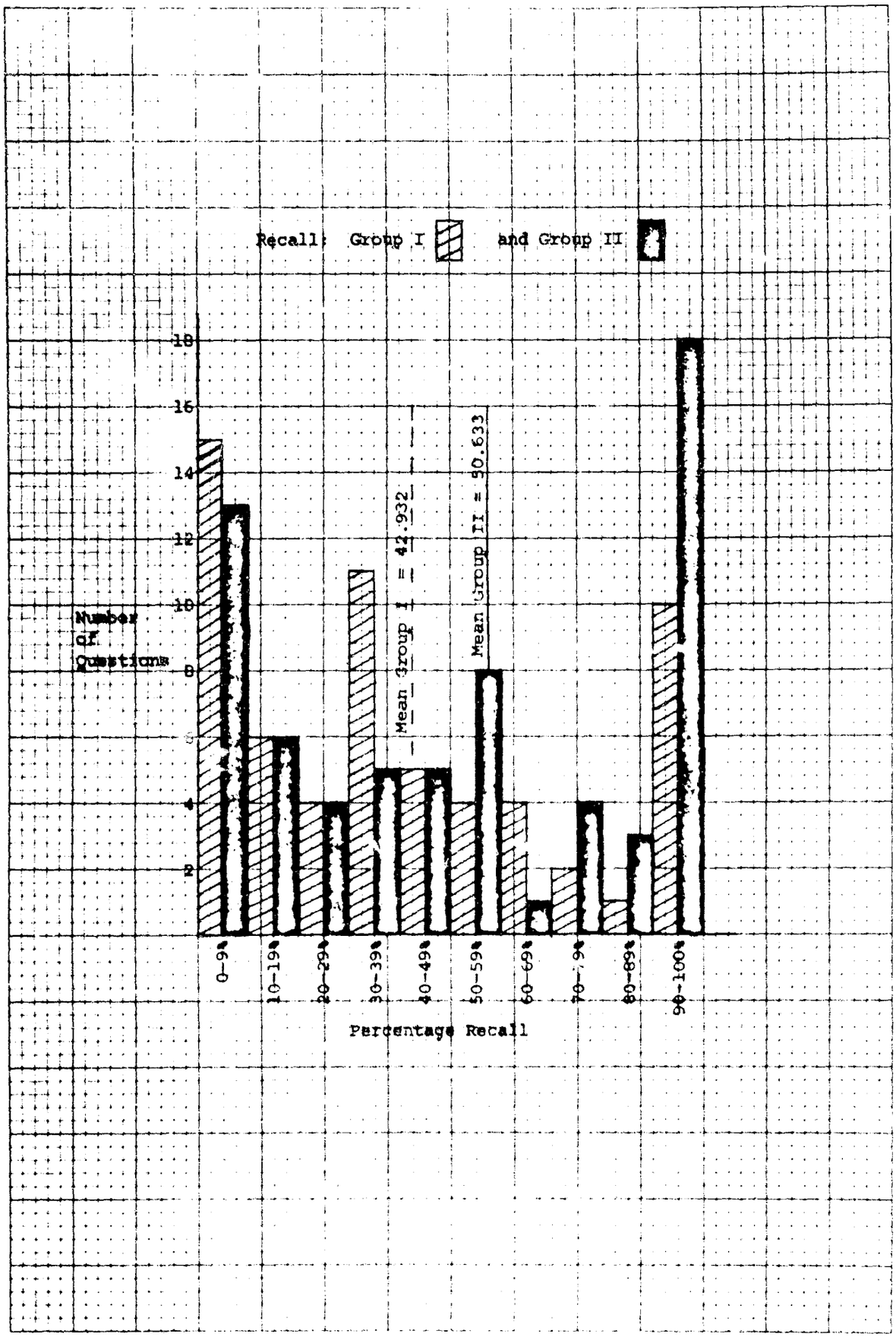
$$\text{Time recall factor} = \frac{\text{Time for searching a tool + (number of false drops x time to remove false drop)}}{\text{Probability of success (recall probability)}}$$

The time for removing a false drop is a variable since it will differ in different indexing systems. The time is dependent, among other things, on the time it takes to locate the document and the time it takes to determine that it is a false drop. We have used a figure of two minutes in our tests of the indexes since it is our estimate that it takes about a minute to locate and refile the document and about a minute to determine that it is a false drop. It should be pointed out that the lower the TRF, the better the index.

III. Search results

a. Group I and Group II; Index A and Index B; Index A and Index E

In this part of the report, comparative search results are given for the two groups of searchers, Group I and Group II, and the index variables tested - the added keywords (Index B) and the single access point per document alphabetic subject index (Index E), both tested against Index A. The results are characterized in terms of recall, precision, search time, TRF, ERF, and E in Figs. 3-20. In each figure the mean search results per question for the groups of searchers as well as the overall mean for the groups of searchers are given. The data in Figs. 3-20 are summarized in Table 3. The search results are analyzed for statistically significant differences with the aid of two tests: the



10 squares to the inch

Fig. 3. Recall - Group I and Group II

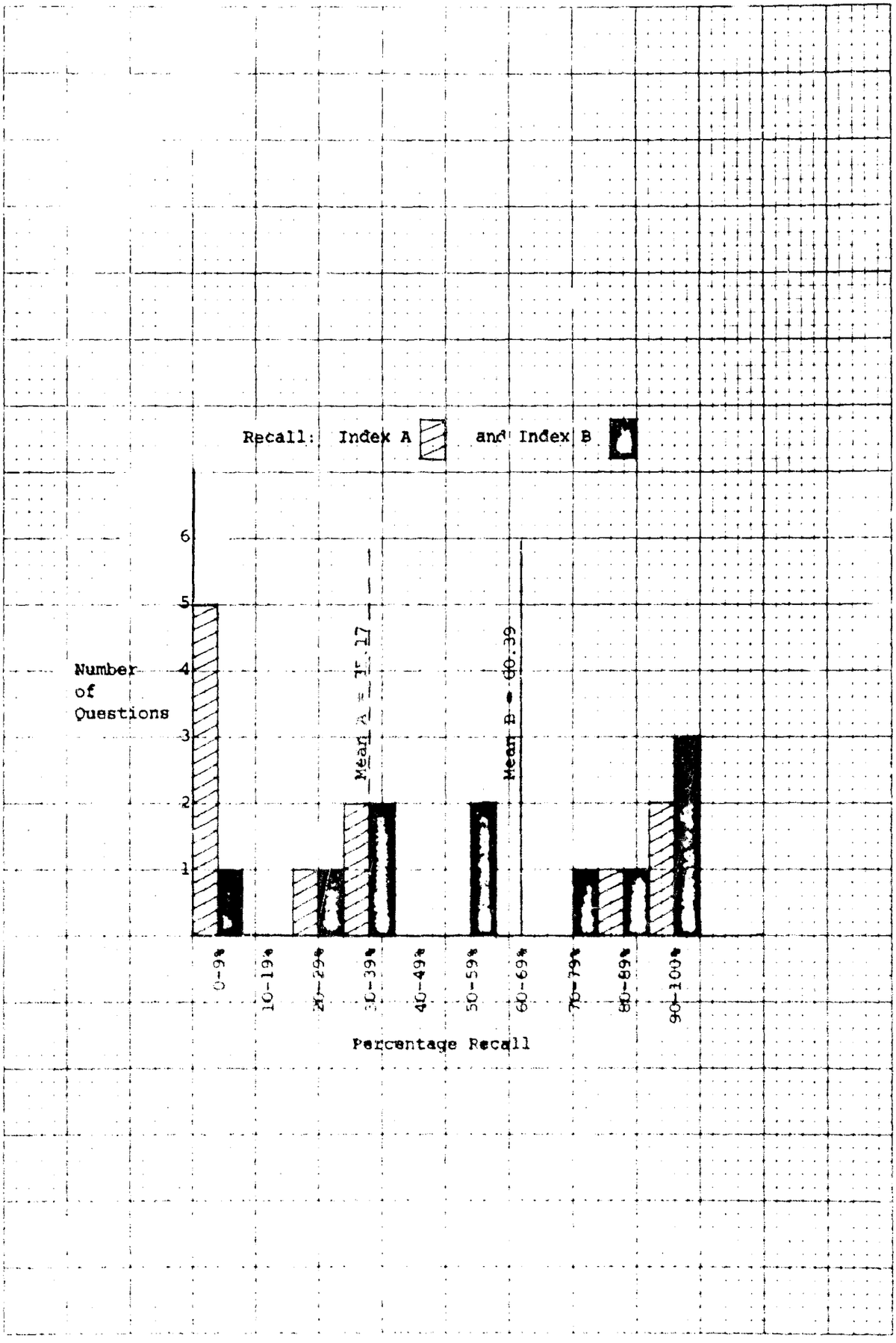
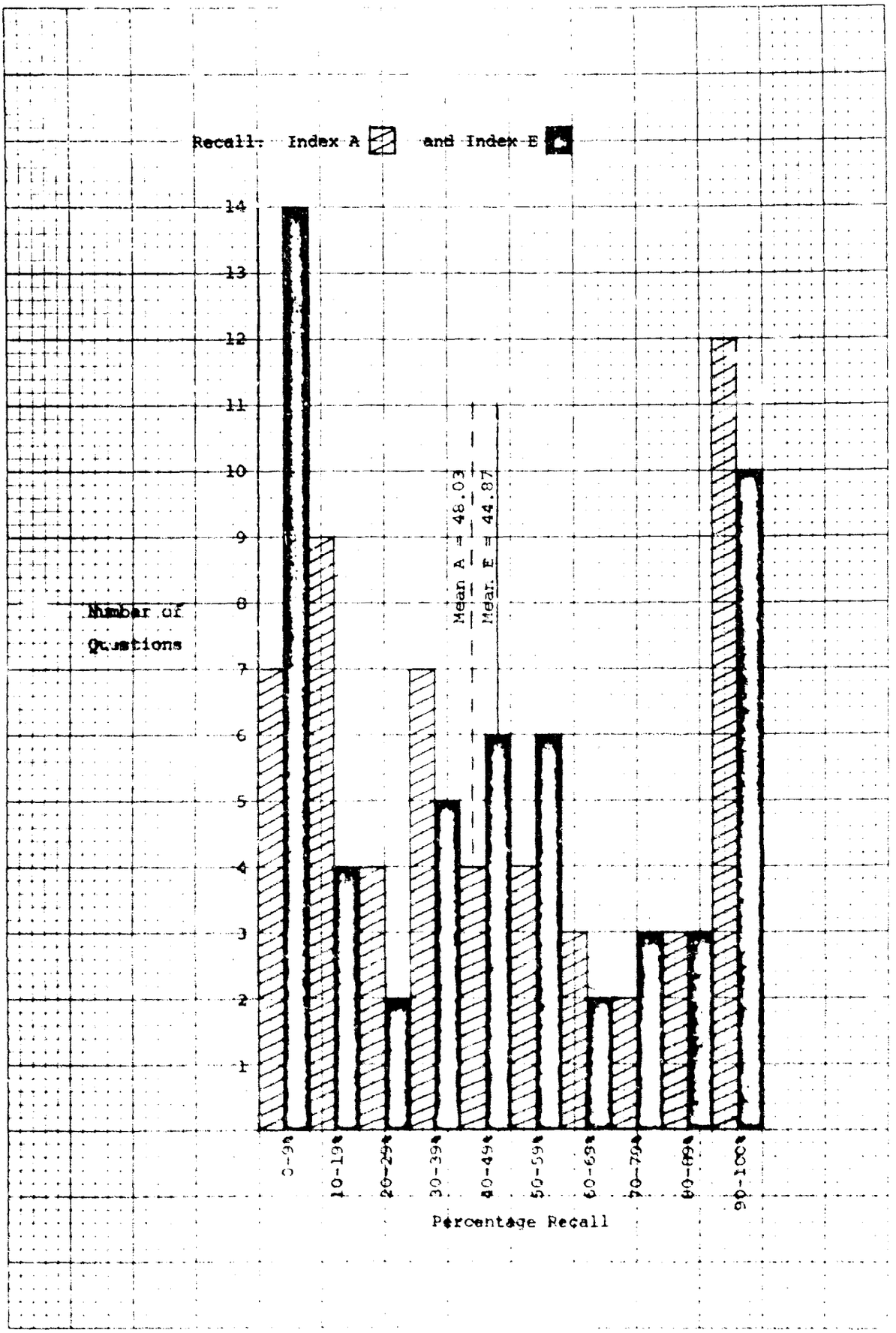
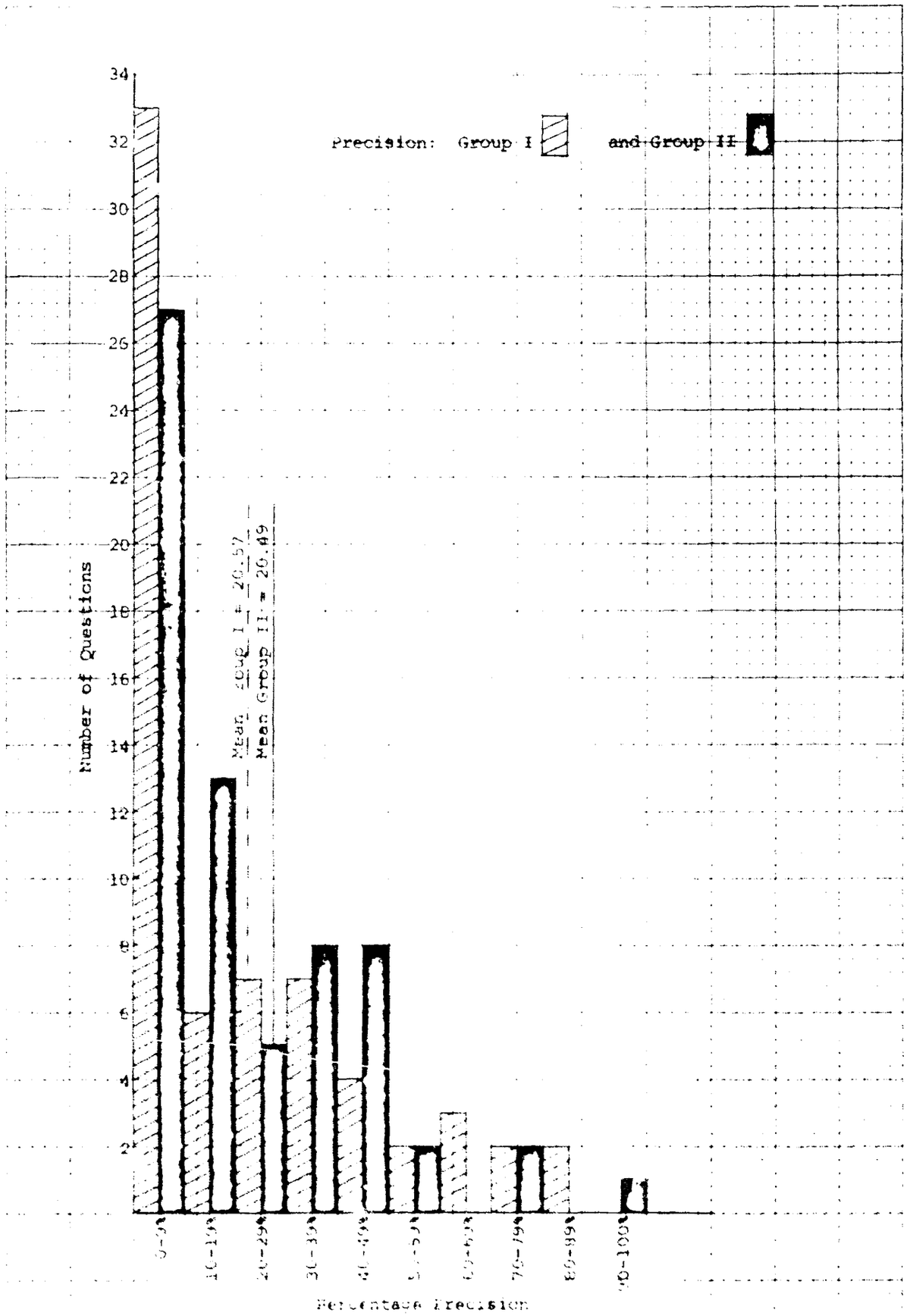


Fig. 4. Recall - Index A and Index B



10 Squares to the Inch

FIG. 5. Recall - Index A and Index E


Precision: Group I and Group II
Number of Questions
Mean Group I = 26.57
Mean Group II = 26.49
Percentage Precision

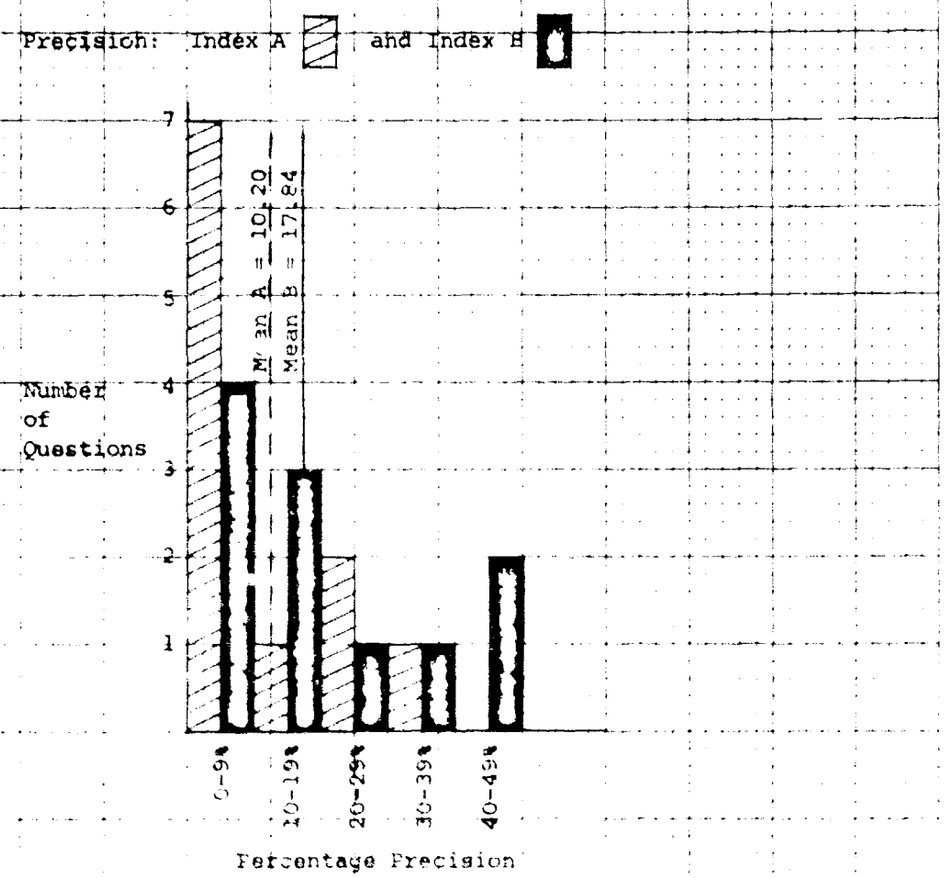


Fig. 7. Index A and Index B - Precision

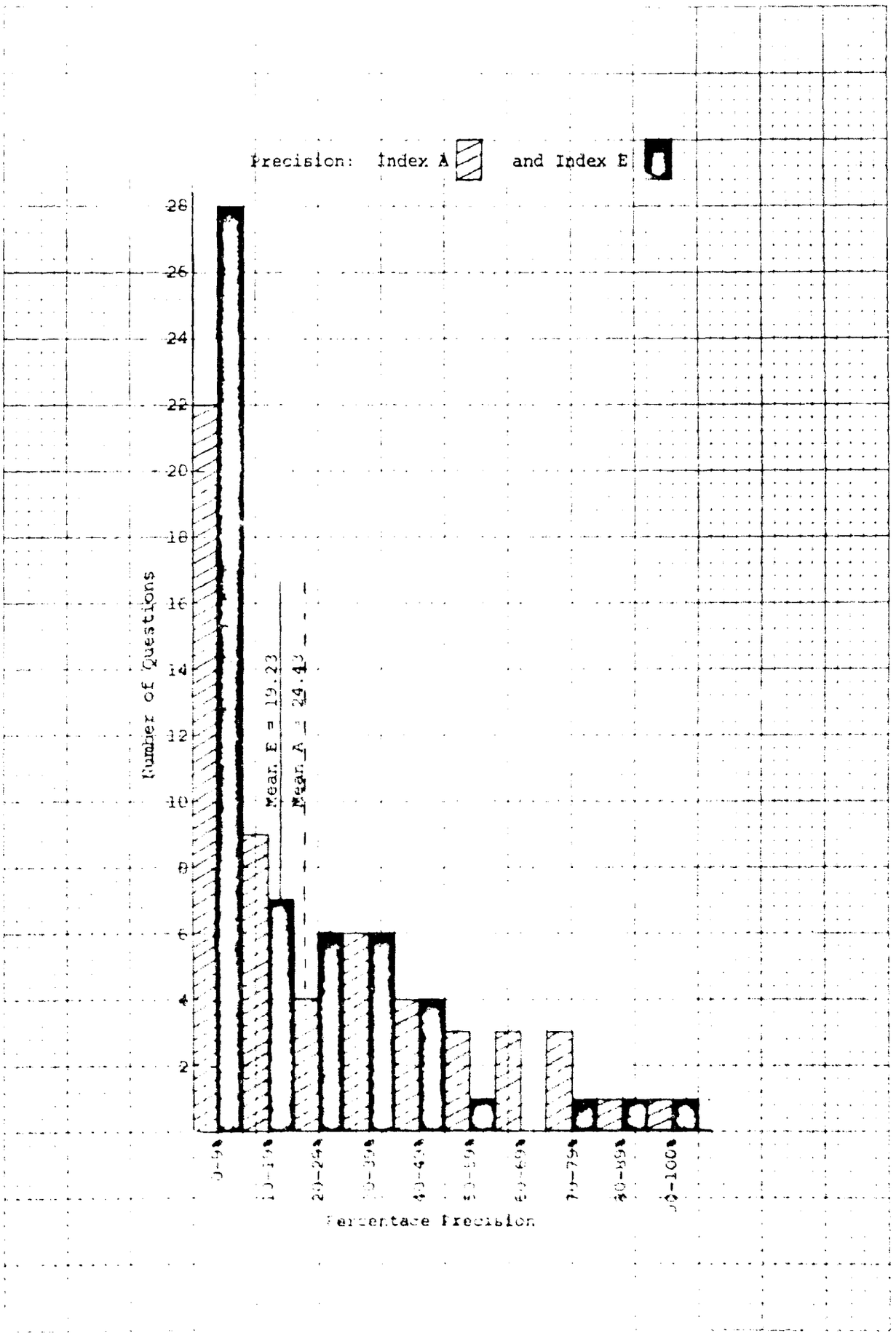


Figure 1. Precision: Index A and Index E

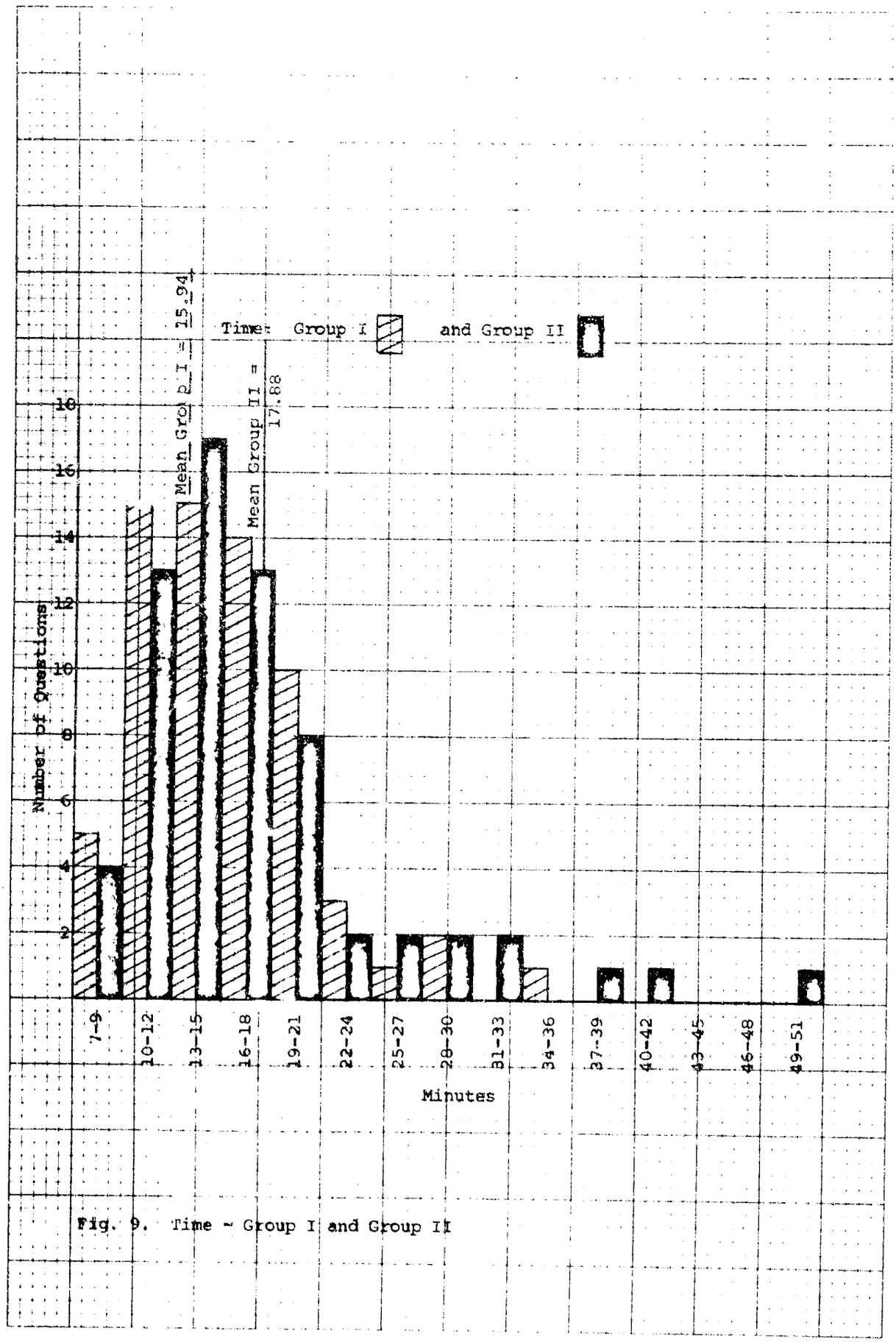


Fig. 9. Time - Group I and Group II

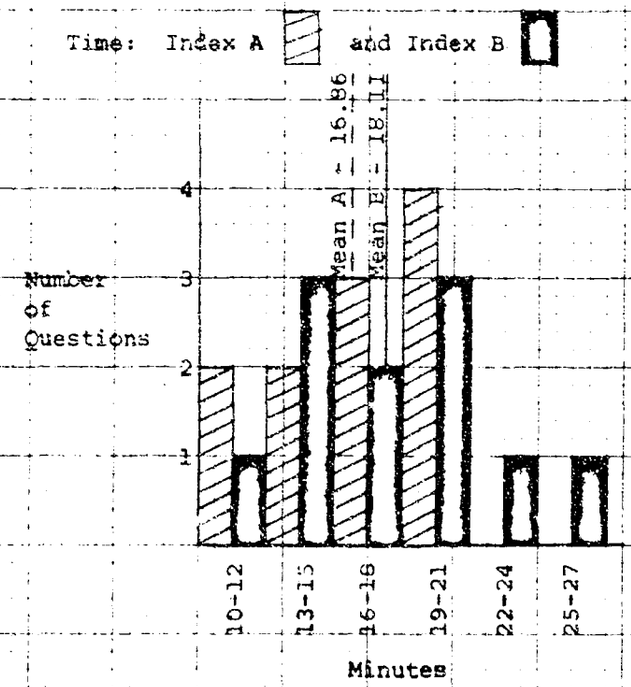


Fig. 10. Time - Index A and Index B

10 Squares to the Inch

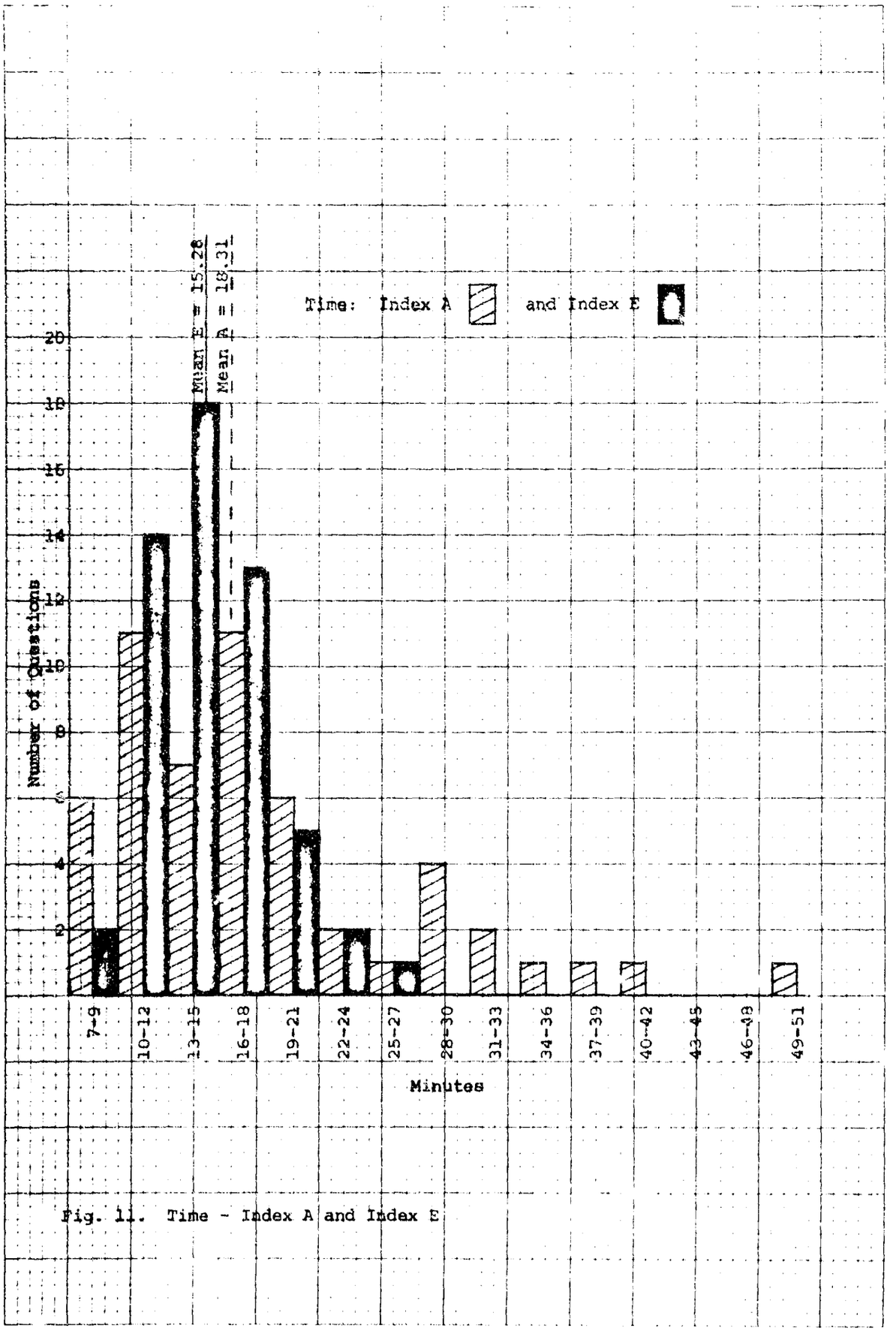


Fig. 11. Time - Index A and Index E

10 Squares to the Inch

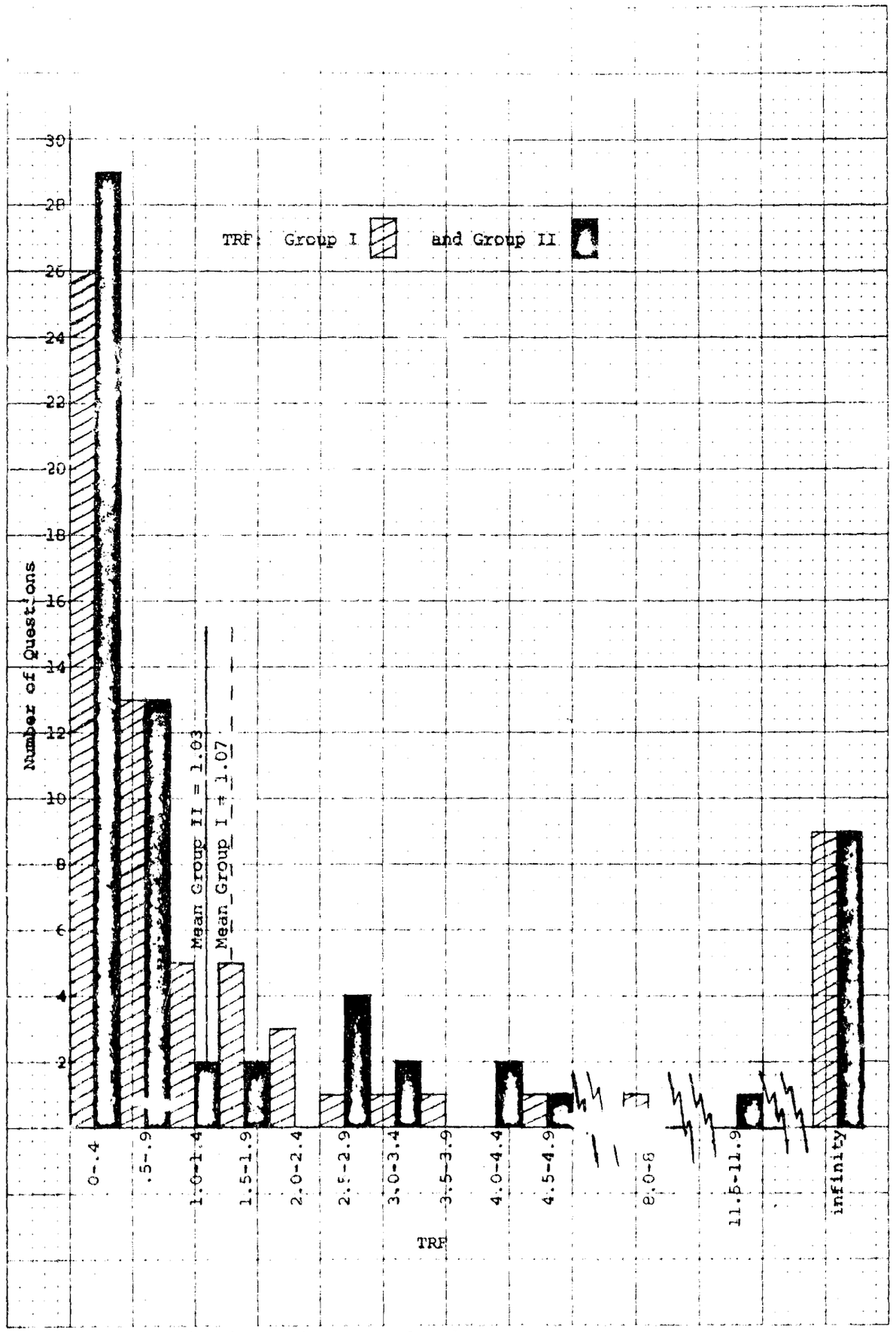


Fig. 12. TRF - Group I and Group II

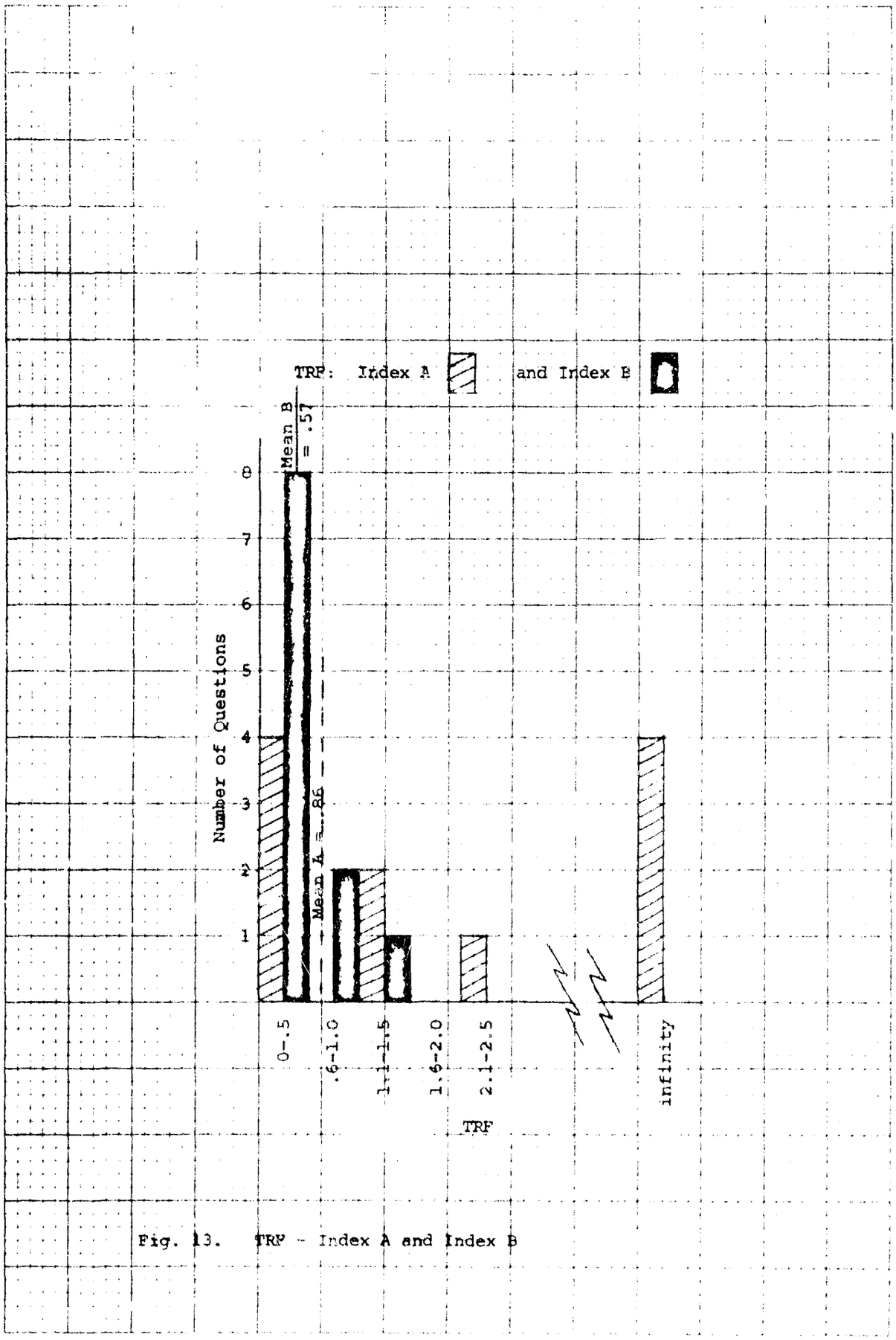


Fig. 13. TRF - Index A and Index B

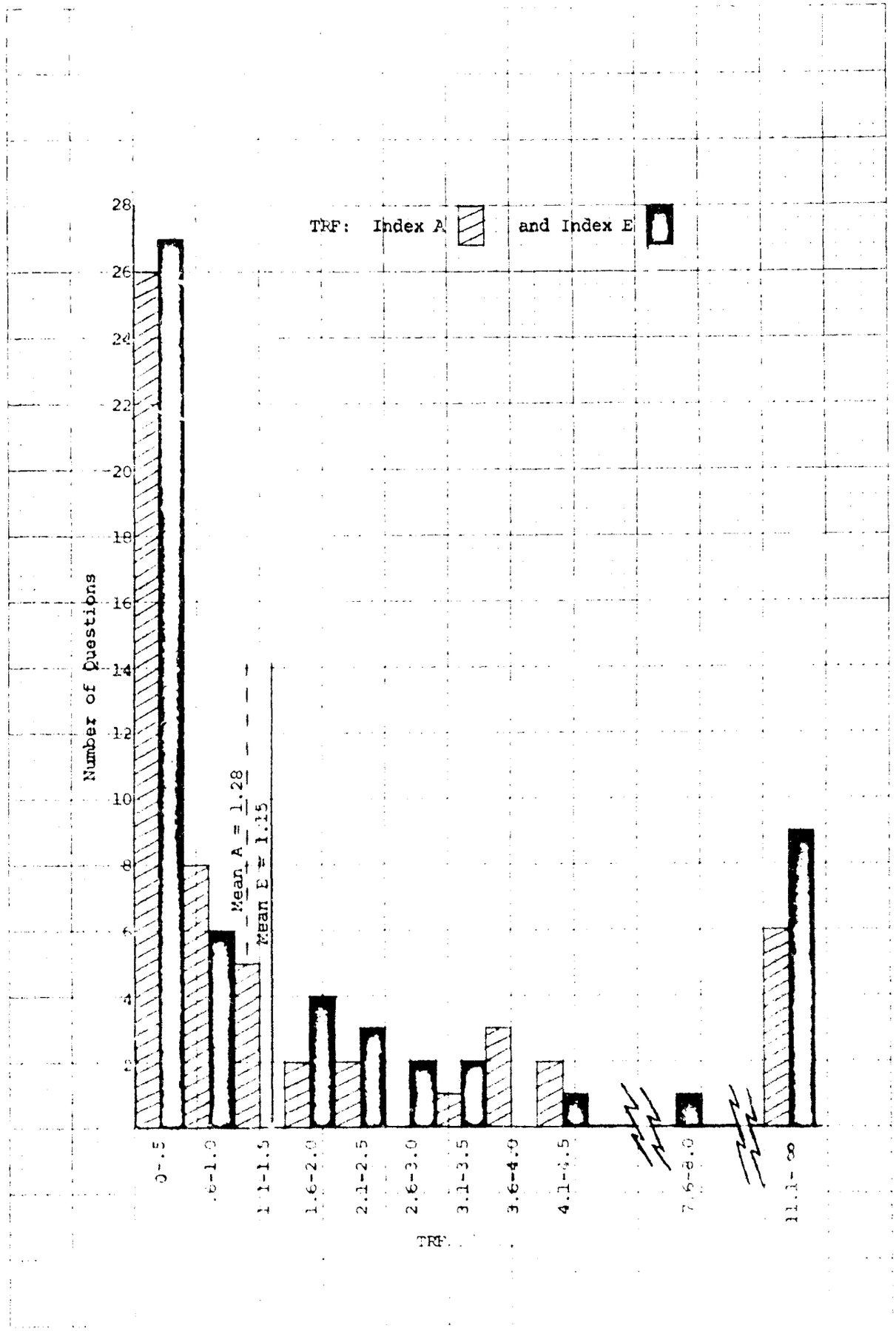


Fig. 14. TRF - Index A and Index E

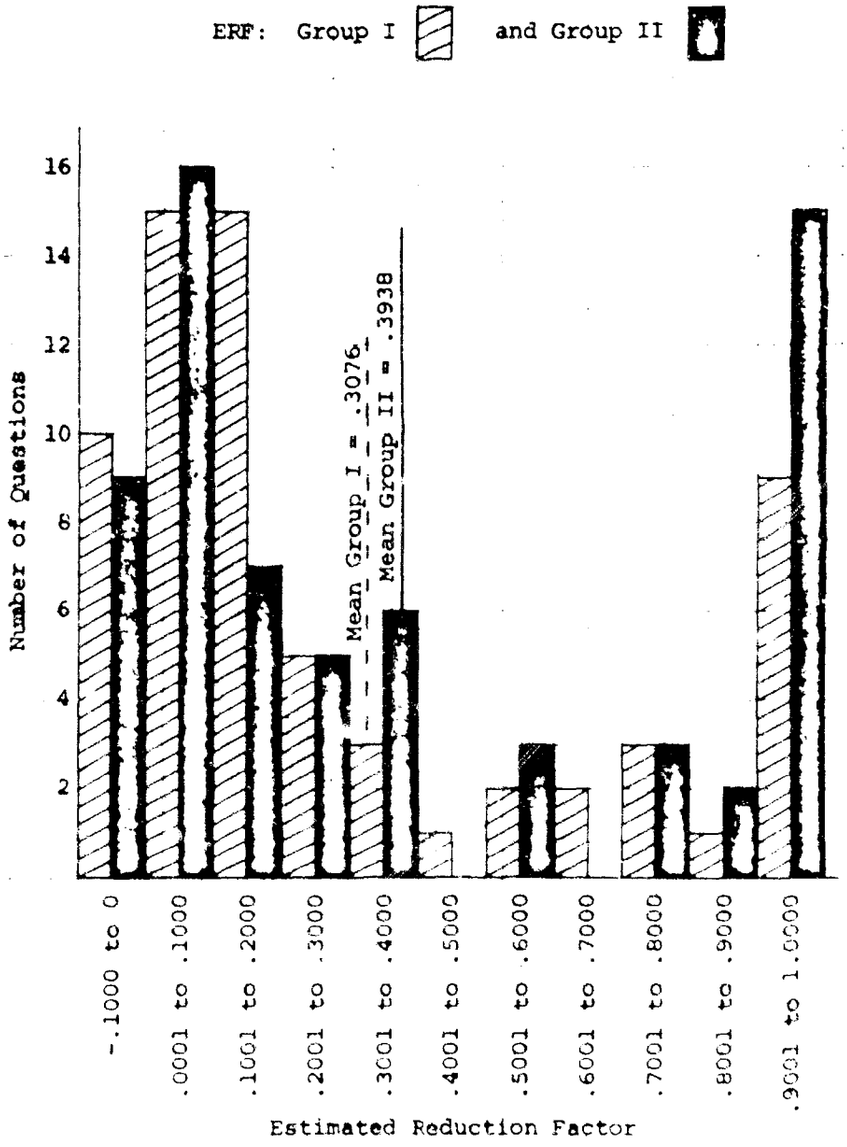


Fig. 15. ERF - Group I and Group II

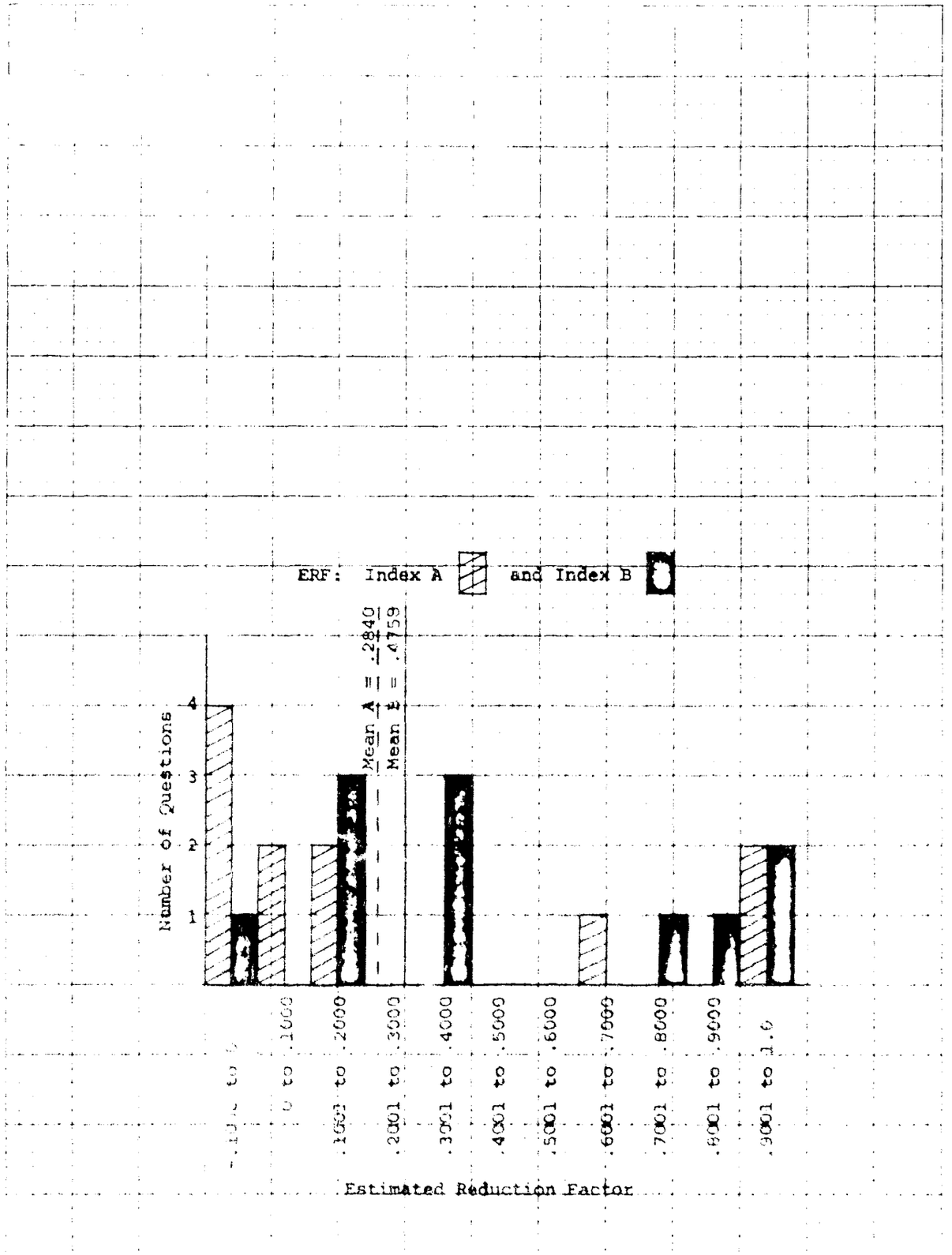
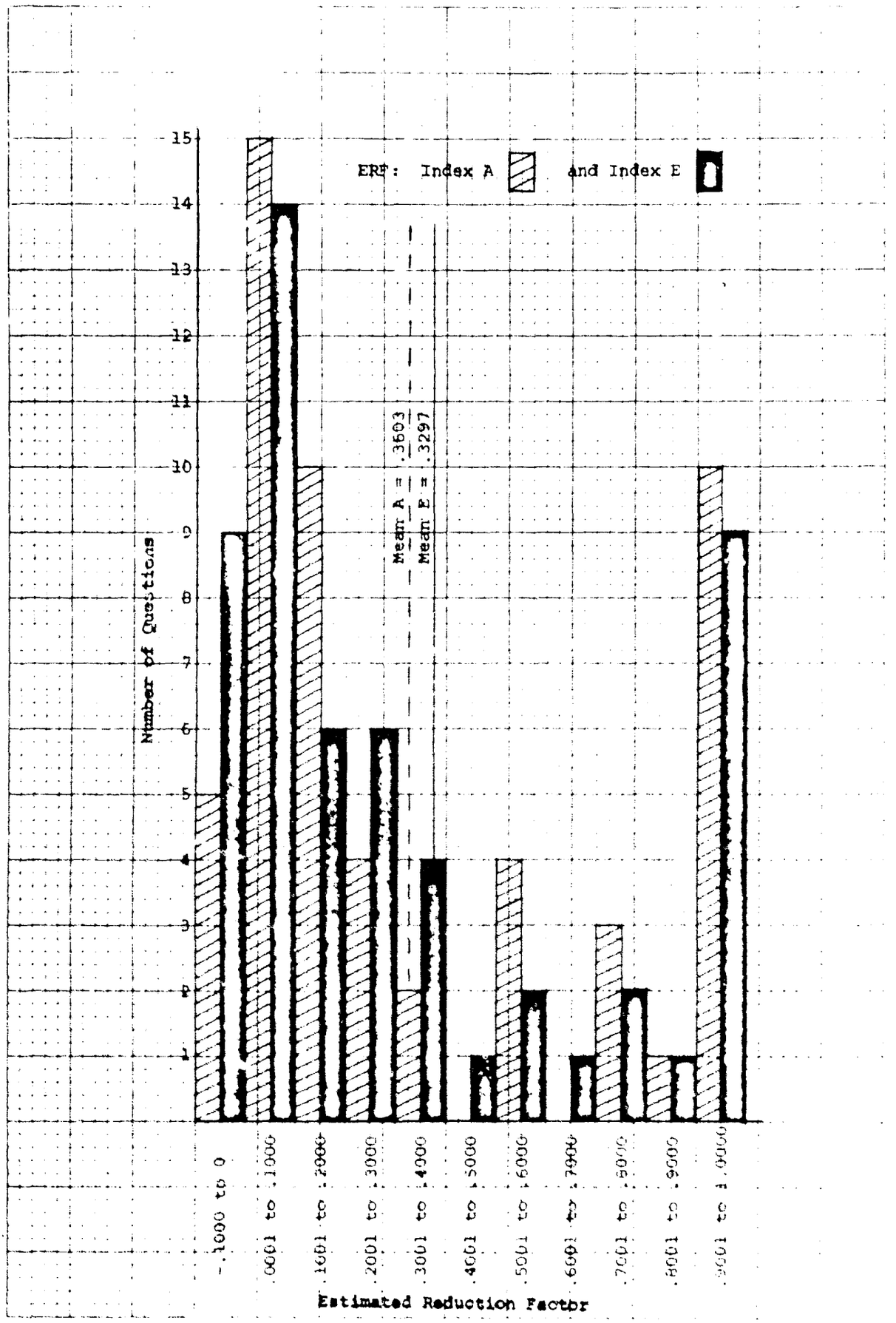


Fig. 16. ERF - Index A and Index B



10 Squares to the Inch

Fig. 17. ERF - Index A and Index E

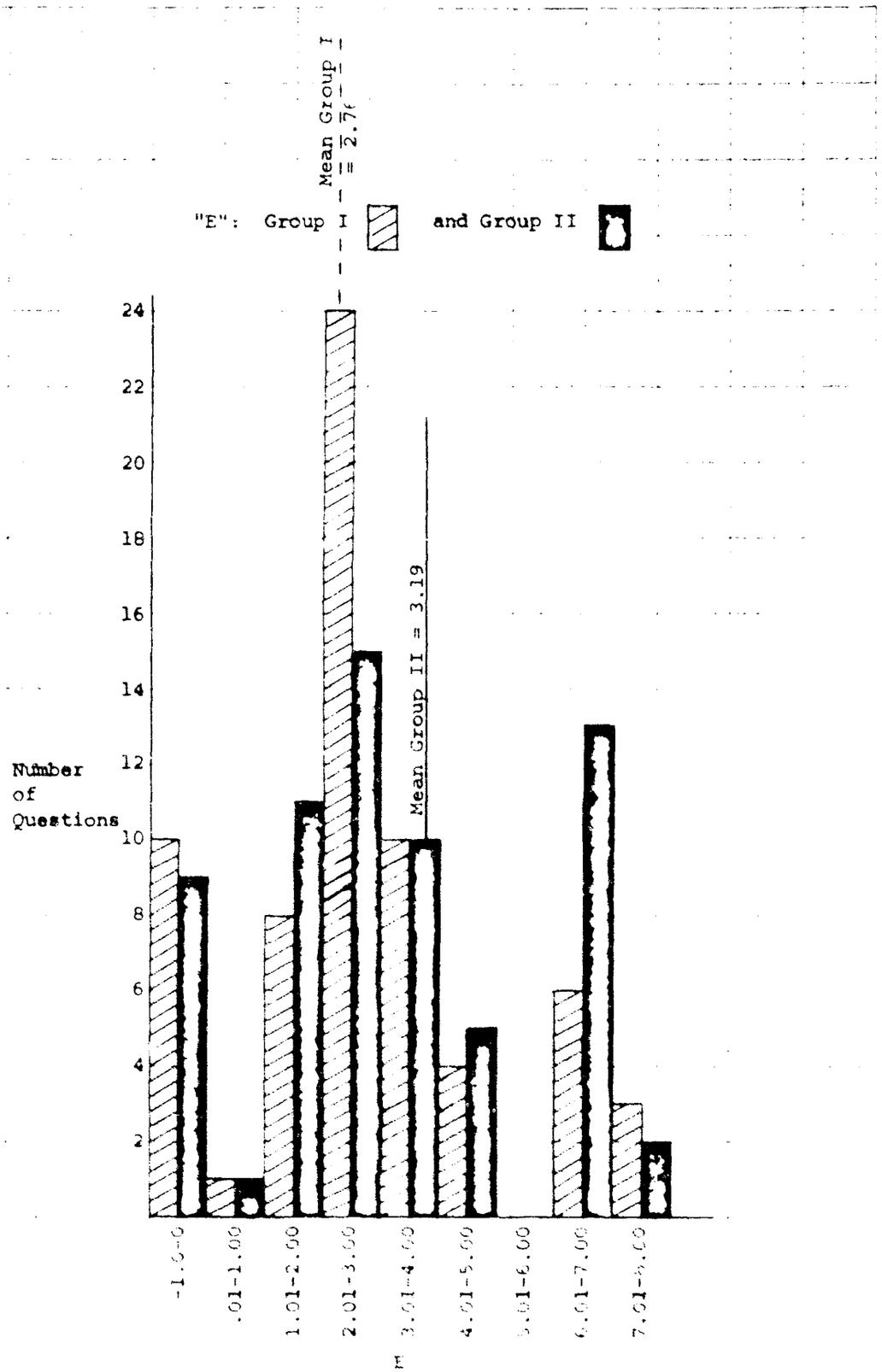


FIG. 18. E - Group I and Group II

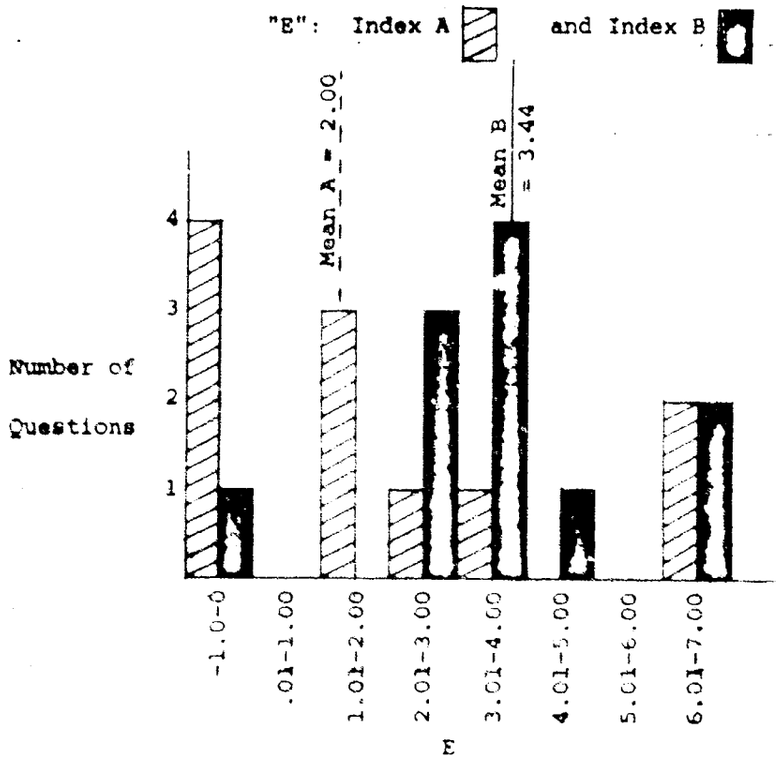


Fig. 19. E - Index A and Index B

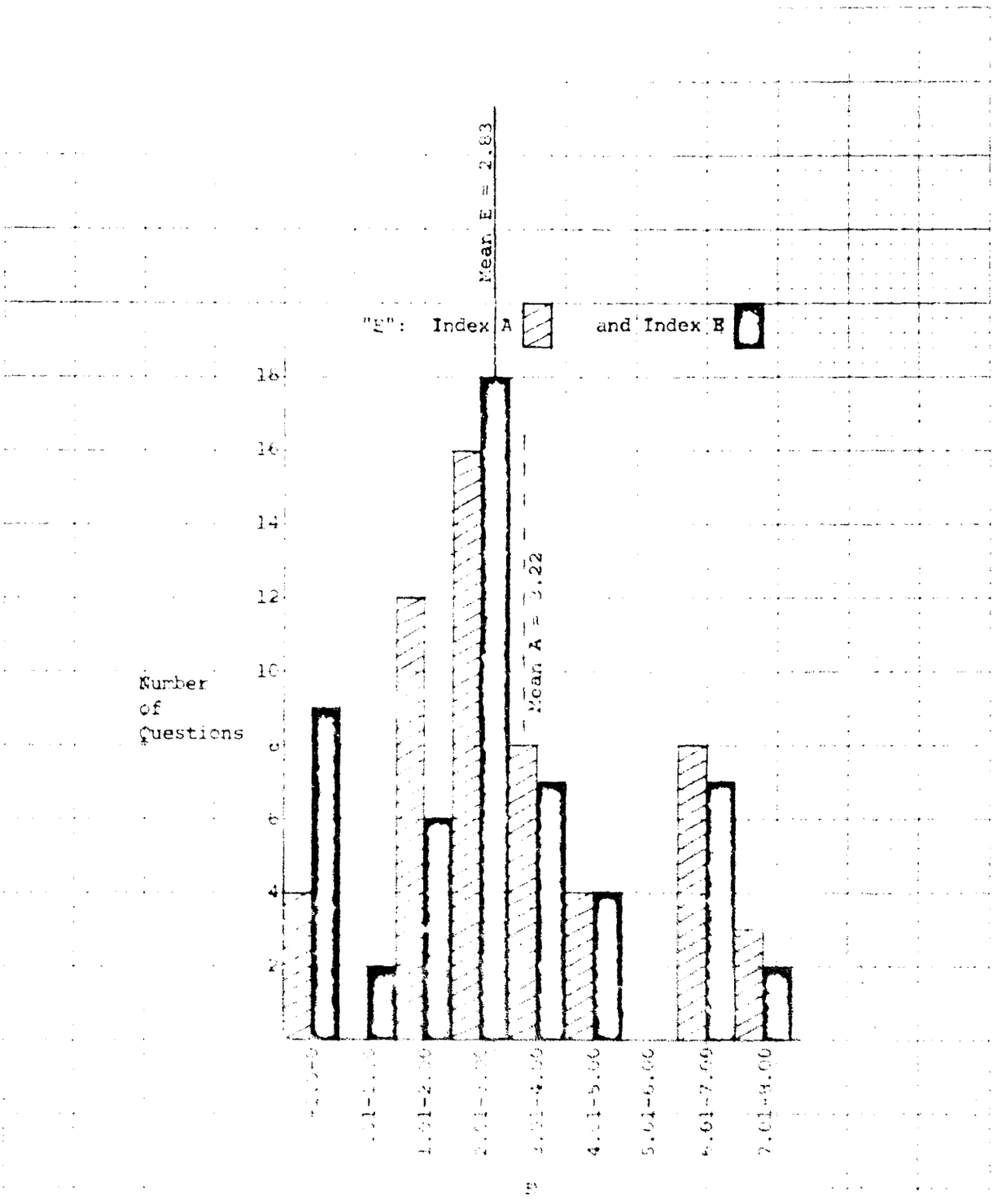


Figure ... E - Index A and Index B

Table 3

Summary of Search Results, Groups I and II; Indexes A and B; Indexes A and E

	Group I	Group II	Means diff.	A	B	Means diff.	A	E	Means diff.
Recall	42.93	50.63	7.70	35.17	60.39	25.22	48.03	44.87	3.16
Precision	20.57	20.49	.08	10.20	17.84	7.64	24.43	19.23	5.20
Time	15.94	17.88	1.94	16.86	18.11	1.25	18.31	15.28	3.03
TRF	1.07 (57 of 66)	1.03 (56 of 66)	.04	.86 (7 of 11)	.57 (11 of 11)	.29	1.28 (50 of 55)	1.15 (46 of 55)	.13
ERF	.3076	.3938	.0862	.2840	.4758	.1918	.3603	.3297	.0306
E	2.76	3.19	.43	2.00	3.44	1.44	3.22	2.83	.39

analysis of variance and the signed pair tests. The signed pair test was also used by Salton in his analysis of search results obtained in his machine prepared and searched indexes.[7] The analysis of variance tests were run with the aid of a computer program.[8] Results of these tests of statistically significant difference are given in Table 4.

b. Index C and Index D

Neither Index C, the index with only 16 delete words, nor Index D, the index with elements of vocabulary control, were actually prepared in their final form. These indexes were therefore not searched by the graduate chemistry students. The anticipated size of Index C and computer program problems in the preparation of Index D discouraged us from preparing these two indexes. Instead, simulated rather than actual searches were made with these indexes.

For Index C, titles of relevant documents were checked against the corresponding questions to determine whether any of the words deleted in Index A but not deleted in Index C would have retrieved the relevant documents. Five such delete words from Index A were located. Table 5 lists these five delete words along with the number of relevant documents that might have been retrieved with each word. The total number of relevant documents for each of the questions for which the delete word might have been useful is given in parentheses.

For Index D, titles of relevant documents were checked against the corresponding questions to determine how many additional relevant documents might have been retrieved with the aid of the cross-references that were to be provided in this index. The 12 cross-references in Index D that might have been of assistance in selecting additional relevant

Table 4

Results of Tests of Statistical Difference

	Analysis of Variance 0.05 level	Signed-Pair Test 0.05 level
Group I vs. Group II		
Recall	NS	NS
Precision	NS	NS
Time	NS	NS
ERF	NS	NS
TRF	NS	NS
E	NS	NS
Index A vs. Index B		
Recall	NS	S (B)
Precision	NS	S (B)
Time	NS	NS
ERF	NS	NS
TRF	Not run	NS
E	NS	NS
Index A vs. Index E		
Recall	NS	NS
Precision	NS	NS
Time	S (E)	NS
ERF	NS	NS
TRF	NS	NS
E	NS	NS

NS = not significant
S = significant
Symbol in parentheses indicates which index or group is better.

documents are listed in Table 6. This table lists the cross-references and the number of additional relevant documents that might have been

Table 5

Index A - delete words that might have been used to retrieve relevant documents

Delete word	Number of documents that might have been retrieved
theory	1 (of 16 relevant)
forming	1 (of 2 relevant)
studies	1 (of 12 relevant)
element	2 (of 11 relevant)
elements	1 (of 20 relevant)

retrieved with the aid of these cross-references for 12 out of the 66 questions. It should, however, be pointed out that the searchers did retrieve some of these documents with Index A, an index without cross-references. From this we can conclude that in at least some of the cases, the searchers directed themselves to the appropriate words without the benefit of cross-references.

IV. Discussion of results

a. Index A and Index B, and Index A and Index E

No statistically significant difference was found with either the analysis of variance or signed pair test for the six sets of search results obtained by the Group I and Group II searchers. We can therefore conclude that any difference between indexes is not attributable to the difference between the groups of searchers.

No significant difference was found with the analysis of variance test for the five sets of search results for Index A and Index B that

Table 6

Cross-references for Index D that might have been used to retrieve relevant documents

Keyword	Cross-reference	Additional relevant documents
lanthanides	rare-earths	2
lanthanide-sulfate	rare-earth-sulfates	1
heat-data	thermodynamics	1
lanthanides	rare-earths	1
lanthanide	rare-earth	2
lanthanides	rare-earths	7
spectroscopy	spectra	2
rare-earths	lanthanides	1
lanthanide	rare-earth	1
lanthanides	lanthanons	1
ion-exchange	cation-exchange	5
ion-exchange	anion-exchange	2
resin	polymer	2
structure-of-water	water-structure	6
structure-of-water	water-structure	2

were compared. No analysis of variance was run on the TRF results since only seven out of 11 pairs of search results could be compared. Reasons for this will be explained in the discussion of the TRF measure. There is a statistically significant difference at the 0.05 level in the signed

pair test for both recall and precision. Eight and nine out of 11 pairs, respectively search results for recall and precision show Index B to be better than Index A. There is no statistically significant difference at the 0.01 level for either recall or precision between Index A and Index B.

In the comparison of Indexes A and E, only the search time for Index E is significantly better at the 0.05 level with the analysis of variance test. This difference does not hold at the 0.01 level. All the other search results for Index A and E for both analysis of variance and signed pair tests show no statistically significant difference between these two indexes.

Even though the statistical tests of search results are in agreement in 14 out of 17 cases, some caution in drawing conclusions from these results is in order. The analysis of variance test assumes normal distribution of search results, something that was not obtained as can be seen from Figs. 3-20. The non-normal distribution of search results is particularly apparent in the recall results. This is due at least in part to the large number of searches that had only one relevant document. When this was the case, either 100% recall was achieved when the relevant document was retrieved or 0% recall was achieved when the relevant document was not retrieved.

The signed pair test is based on any difference between pairs of search results whether this difference is a fraction of a percentage point or 20 or more percentage points. There is a question whether a fraction of a percentage point difference is meaningful. Yet we cannot set the minimum meaningful difference at a specified quantity, partly because we do not know at this time what this quantity should be and

partly because if we did do this we would run the risk of introducing a bias in the test.

We anticipated Index A to perform better than Index E since Index A had more access points per document than Index E. The following reasons might explain why Index A did not perform significantly better than Index E.

* The additional information per alphabetic subject index entry as compared with the KWOC index entry, the subject heading, may have offset the disadvantage of only a single access point per document. This might have been the case for both the selection of relevant documents and the rejection of non-relevant documents.

* There are relatively few subject headings in the index, only 194, and relatively few documents under each subject heading as can be seen from Fig. 21. This may account for Index F's advantage in search time since less time is required to locate the access point in this index than with Index A. Another reason for the better search time for Index E is that fewer access points were searched per question in Index E than in Index A.

Index B had statistically significant better recall and precision than Index A with the signed pair test though not with the analysis of variance test. Index B was expected to show better recall and precision since it included the added keywords chosen by the two chemists for the indexes to their documents. The better recall can be explained by the additional access points per document (a factor that was, however, of questionable value in the Index A versus Index E comparison). The added keywords tagged along with every index entry, whether or not they were the access points. This added information per entry may have been of use in eliminating non-relevant documents and may therefore explain the

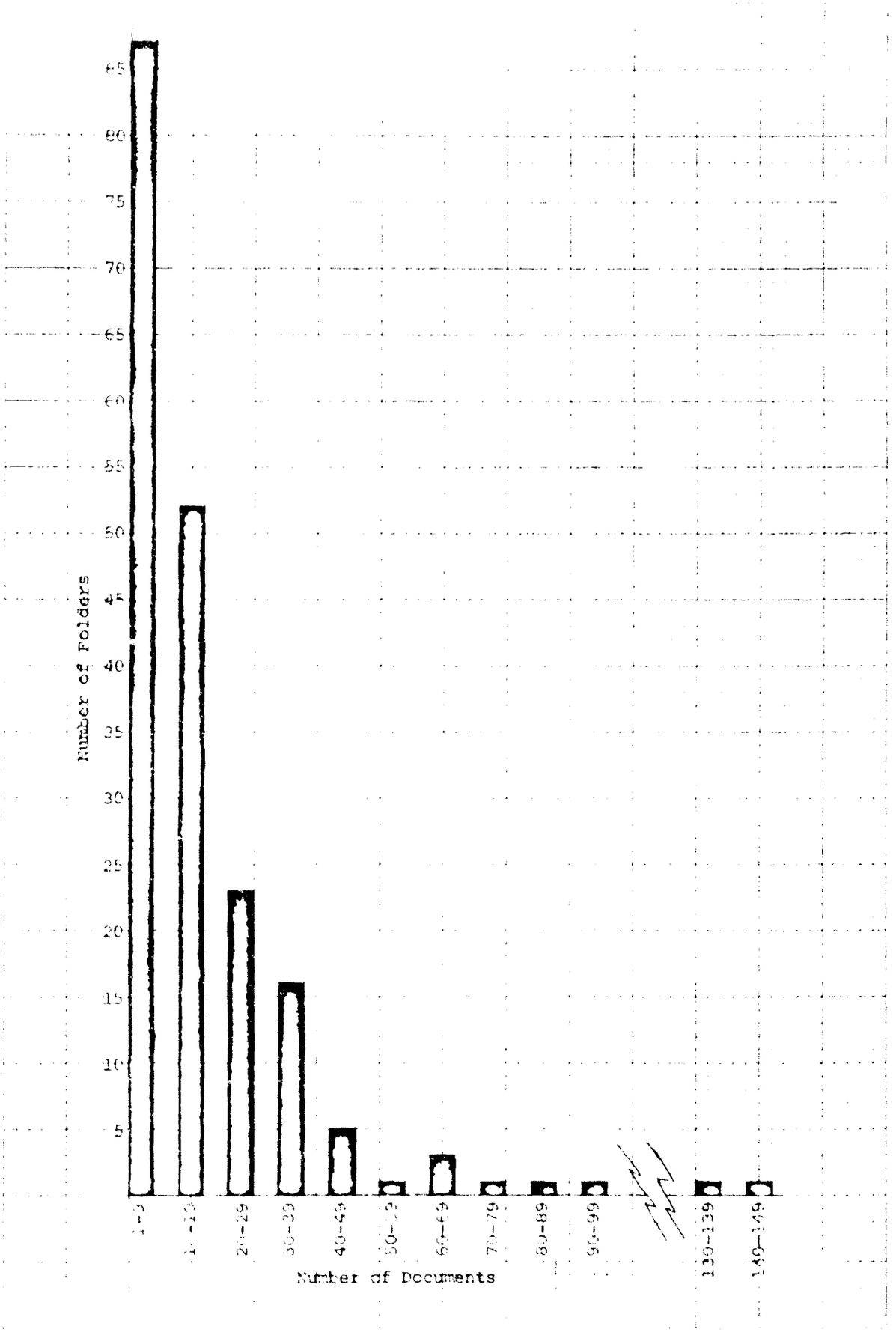


Fig. 21. Number of Documents per Folder

better precision of Index B than Index A. No better (faster) search time was either expected or obtained with Index B since the added keywords would have a tendency to lengthen the search time if they affect search time at all. Forty-eight questions out of the 66 questions came from the two chemists who added keywords to their indexes. Forty-eight questions could, therefore, have been searched in Index B instead of the 11 questions that were actually searched in it. Only 11 questions were searched in Index B since only those questions were searched that were considered to be easier to search with this index. These are the questions for which useful access points could be supplied better or only by means of the added keywords. Had all 48 questions been searched in Index B, chances are that the difference in performance between the index without added keywords and the index with added keywords would have been even smaller.

b. Single figure index measures

Since none of the comparative search results with measures E, ERF, and TRF on the groups of searchers and the three indexes showed any statistically significant difference, no comments can be made on the discriminating power of the three index measures. In looking at Table 3 we can see that the E and ERF measures are consistently higher (better) for the two groups of searchers and the two pairs of indexes tested. The TRF measure was also consistently lower (better) for the two groups of searchers and for Index A versus Index B. The TRF was better for Index B rather than Index A as was the case for both the E and ERF measures. This result is probably attributable to the better search time obtained with Index B.

There were some problems in using the single figure index measures and these problems will now be discussed. An adjusted value of 99.99% recall instead of 100% had to be used for calculating E values for search results with 100% recall. Since the percentage of false drops was quite low in most cases, a graph paper other than the one originally suggested by Swets had to be used. Dr. Swets kindly supplied a different graph paper, K-E 47-8062, that could be used. It was not possible to determine the slope for each search both because a large number of questions yielded only one relevant document and because of the small number of searches per question. Instead, average figures were used for number of relevant documents retrieved, number of relevant documents not retrieved, number of non-relevant documents retrieved, and number of non-relevant documents not retrieved. The measure E was derived by the use of the operating characteristic curve published in the Swets paper.[9]

For questions that yielded only one relevant document and one that was not retrieved, a negative value was obtained with the ERF measure. This means that for this particular type of search, the search time with an index takes longer than purely random retrieval. This may or may not be so.

The TRF measure could not be used with 0% recall search results. Such search results yielded an unrealistic TRF value of infinity. Search results with 0% recall were therefore omitted from the statistical analysis, hence the smaller number of paired search results that were compared for the TRF measure than for other search measures. To get around this problem, the recall at both extremes, 0% and 100%, might be adjusted upward and downward, respectively. The two minute time for identifying and rejecting a false drop may or may not be realistic. This figure can be adjusted to

represent the time for any particular indexing system.

c. Index C and Index D

The KWOC index that was to have only 16 delete words instead of the 628 delete words would have been a very large index indeed, several times larger than the index with 628 delete words. In this instance at least, one does not get much from the extra bulk. Only six relevant documents that were not retrieved with Index A might have been retrieved with the minimum delete word index. Furthermore, it is questionable whether a searcher would look under two of the words that might have retrieved the additional relevant documents, the words "forming" and "studies".

An examination of the search results obtained with Index A, the index without elements of vocabulary control, indicated that the searchers of this index thought of and used some of the cross-references that were to be provided in Index D. The extent of this generation of cross-references by the searchers will be determined from a further analysis of the search results.

d. Some general conclusions and suggestions for future work

To anyone acquainted with index testing, our general conclusion will not come as a surprise: there are problems in index testing with which we have not as yet learned to cope. Some of these problems, along with comments on what, if anything, was done about them in our tests of the indexes, will now be summarized.

1. The subjectivity and inconsistency of relevance judgments have been problems in previous tests of indexes, and these problems have not been resolved in our tests. Two steps have been taken to reduce the problem of relevance judgment. Relevance judgments by the original searchers, the three chemists who searched their own document collections, were used for the characterization of relevant documents. Searchers with similar educational background, graduate chemistry students, were employed to conduct the test searches. Also, search results by groups of searchers rather than by individual searchers were used in the analysis of the test results.

2. There is a subtle but important distinction between a statistically significant difference and a meaningful difference in the performance of two indexing systems. Statistics enable us to determine whether a difference is more than a matter of chance. But statistically significant differences may or may not be meaningful differences from the point of view of the overall performance of the indexing systems. This question will remain unresolved until we test the overall indexing system rather than the output of the system and until we can specify what is a meaningful difference between indexing systems.

3. Search results for questions with only one relevant docu-

ment give extreme recall values of 0% when the document is not retrieved and 100% when the document is retrieved. These single relevant document search results may carry an unduly large weight in averaging search results. Search results for searches with one relevant document also caused difficulties in calculating each of the three single figure measures of index performance. The number of relevant documents per question should perhaps be one of the index variables to be tested. We plan to analyze separately the search results for the 11 questions that had only one relevant document.

4. All other things being equal, the cost of preparation of an index with elements of vocabulary control is greater than the cost of preparation of an index without elements of vocabulary control. Again, with all other things being equal, an index with elements of vocabulary control is likely to give higher recall than an index without elements of vocabulary control. However, some evidence has been collected in the course of this study that the searchers will think of and supply their own cross-references in an index without elements of vocabulary control. Further work is therefore suggested to determine both the cost and the benefit of elements of vocabulary control in an indexing system.

There is an indication that for a collection of about 3200 documents a single access point per document alphabetic subject index is as good as a multiple access point per document keyword from title index. We would like to see these findings tested on indexes that are larger by an order of magnitude. This could be done with the 66 questions that were used in this study by modifying an existing index, Chemical Abstracts Service' Chemical Titles. The magnetic tapes for several issues of this

keyword from title index would be merged and a merged index to about 30,000 documents would be printed. The resulting keyword from title index would then be compared with a single access point per document index to the same group of documents. The single access point per document index would be the listing of the document citations under the sections headings used in Chemical Abstracts for grouping the abstracts of the documents.

V. Bibliography

1. Jahoda, G., Hutchins, R. D. and Miller, D. M. Analysis of case histories of personal index use. Proc. 1966 Meeting of the American Documentation Institute, 1966, pp. 245-54.

2. Luhn, H. P. Keyword-in-context index for technical literature. Am. Doc. 11:291 (1960).

3. Fairthorne, R. A. Basic parameters of retrieval tests. Proc. 1964 American Documentation Institute, vol. 1, 1964, pp. 343-45.

4. Swets, J. A. Information retrieval systems. Science 141:245-50 (1963).

5. Cooper, W. S. Expected search length: a single measure of retrieval effectiveness based on the weak ordering action of retrieval systems. Am. Doc. 19:30-41 (1968).

6. Lazorick, G. J. and Minder, T. L. A least cost searching sequence. Coll. Res. Lib. 25:126-28 (1964).

7. Salton, G. Automatic information organization and retrieval. McGraw Hill, 1968, pp. 310-15.

8. Dixon, W. J., ed. BMD biomedical computer programs. Berkeley, Univ. of California Press, 1968, pp. 586-95.

9. Swets, op. cit.

VI. Acknowledgments

We would like to acknowledge the help of Professor Duane A. Meeter and Mrs. Joan Chellis. Professor Meeter assisted in the use of the statistical tests. Mrs. Chellis was of help in all phases of the study, including the preparation of the figures and tables.

Appendix A

QUESTIONS FOR KWOC INDEX TEST

1. A chromatographic method for purifying lipids.
2. Want an experimental method for preparing methyl esters of fatty acids.
3. Experimental details concerning an extract of penicillium which converts methylsalicylic acid to patulin.
4. Identification of methyl derivative of triacetic acid lactone in penicillium stipitatum.
5. Heterocaryon formation and genetic analysis.
6. What is the content of fatty acids in brain lipids?
7. 6-methyl salicylic acid production in penicillium.
8. Calculate quantity of methylsalicylic acid produced in a cell-free system.
9. Concentration of triacetic acid lactone used to stimulate aromatic biosynthesis.
10. Method of separation of various unsaturated fatty acids.
11. Growth properties of penicillium patulin in relation to methylsalicylic acid biosynthesis.
12. Information on the sulphate complexes of lanthanide elements.
13. Information on the stability constants of lanthanide sulfate complexing.
14. Heat data for rare-earth acetates.
15. Method of preparation of anhydrous salts in connection with spectrophotometric studies of lanthanides.
16. Information on the theory for intensity of lanthanide spectral lines.

17. Spectroscopy of lanthanides.
18. Stability data on chelate complexes of rare-earths.
19. How are the hydrates involved in the method of preparation of volatile complexes of lanthanide elements?
20. Information on stability of volatile complexes of the lanthanide elements.
21. Information on the thermodynamics of water absorption in order to start writing a paper on nuclear magnetic resonance studies of water in ion exchange resins.
22. What is the structure of water?
23. Does the resin matrix cause the structure of water to increase?
24. Method for preparation of diazoethane.
25. Chemical synthesis of phospholipid derivatives.
26. Methods of preparing carbonyls in aqueous solutions in order to get references.
27. The use of borontrifluoromethonyl solutions for transesterifying lipids.
28. Use of fructose di-phosphate in stabilizing the fatty acid synthetase.
29. What are the experimental details in the use of dithiothreitol in the protection of enzymes?
30. References to article that uses dithiothreitol in stabilizing the pigeon liver fatty acid synthetase.
31. Conditions for forming acetyl derivatives of acetyl carrier protein.
32. Reference to a paper which would have a synthesis of acetoacetyl coenzyme-A.
33. Metal-metal bond energy of dimanganese-carbonyl.
34. Stability properties of triacetic acid and its lactones.

35. Method of preparation of 4-hydroxy-6-methyl pyridone and properties.
36. Degradation scheme for the compound patulin in order to test the feasibility of an isotope experiment.
37. Occurrence of hydrocarbons in Sarcina lutea.
38. Absorption spectra of dithiothreitol.
39. Method for making trimethylsilyl derivatives of sugars.
40. Quantity of patulin produced in cell free system.
41. Analysis of fatty acids of marine organisms.
42. Want information on the substitution of phosphorus-trifluoride into cobalt-tricarbonyl-nitrosyl.
43. Quantity of alternariol produced in a cell-free system.
44. Medium for growing Alternaria tenuis.
45. Experimental procedure used in purifying an enzyme from alternariol.
46. Wanted paper chromatography system to analyze triacetic acid.
47. Retention time of hydroxy fatty acids on gas-liquid chromatography.
48. Concentration of methylsalicylic acid which shows inhibitory effects on methylsalicylic acid synthesis.
49. Method of isolation of penicillium patulin metabolites.
50. Conditions used in studying inhibition of enzymes by sulfyhydryl reagents.
51. The tate of water in ion exchange resins.
52. The cumulation of side-products in fatty acid synthesis.
53. Identification of a side-product in fatty acid synthesis.
54. Information on the fission cross-section as a function of mass-number.
55. Conditions for CO_2 malonyl coenzyme-A exchange reaction.
56. Experimental procedure for enzymatic assay.

57. Empirical Zp functions in connection with charge distribution studies on fission.

58. Nuclear magnetic resonance studies on ores, minerals and clays.

59. Information for preparation of a computer program on stability constants.

60. Need information for a talk on new element research.

61. Experimental procedures for the Kuhn-Roth oxidation wanted.

62. Conditions for obtaining the crude fatty acid synthesis from E. coli.

63. Method for terminal methyl group determination by chromic acid oxidation.

64. Effect of cunionine on synthesis of orsellinic acid.

65. Need method for synthesis of orsellinic acid.

66. Stimulation of metabolite production in penicillium urticae barnier by triacetic acid lactone.

Appendix I

KWOC INDEX TEST INSTRUCTION:

A. Indexes A and B

1. Introduction

The objective of this test is to determine how useful a KWOC (Keyword-Out-Of-Context) index is in the locating of all documents in a collection which are relevant to a particular question through the use of keywords in the document titles. A KWOC index is a computer print-out in which significant words from titles, names of authors, and in some cases additional keywords, are arranged alphabetically at the left-hand margin. In the index you will be using, the complete citation and document serial numbers are listed with the alphabetized title words. At the front of the index is a listing of "stopwords" which are words not considered significant and not indexed as keywords.

The keywords in this index are not always in a strict alphabetic order because of a machine idiosyncracy in the processing. For instance, the hyphen acts as if it were a letter of the alphabet which would follow the letter A. Therefore, vapor-pressure appears after vapors, and F-orbitals appear after all the other words beginning with the letter F.

2. General Instructions

You will be given a list of search questions. Each of these questions will have one or more documents that were considered pertinent to the question by the researcher who originally asked the question. You are asked to select keywords under which to look for potentially pertinent titles and then decide whether the title is pertinent. Your decision must be based on the words in the title. You are to attempt

to locate all titles in the index that are pertinent to the question.

The procedure that you are to follow is illustrated by the sample question on the form attached to these directions.

a. Record the time at the start of the search.

b. Read the sample question.

c. "Fluorescence" is chosen as a keyword.

d. Write this keyword in the space provided in the section titled "Keyword Selected."

e. Scan the alphabetically arranged keywords in the KWOC index to see if "fluorescence" appears as a keyword. You will see that there are 11 titles which contain the word "fluorescence."

f. Read each title and determine if it might answer the question.

g. Several titles under this keyword are considered to be pertinent. Write the serial number of the titles you consider to be pertinent in the space provided. If there had not been any titles considered pertinent, you would have written "none" in the serial number space.

h. Choose the next keyword that you will use to search. Any keyword may be used in the search. It does not need to be in the question, but can be a synonym or a more generic or specific form of a word. It may be necessary to look under various forms of the same word when locating information. For example, "fluorescence-spectra," a specifically related form of "fluorescence," is chosen as a keyword. Write this in the section provided for keywords. Turn in the index to the appropriate keyword section. A total of six titles which contain the word "fluorescence-spectra" are listed. Read each title. The title

with the serial number FV-028 is selected as being pertinent. Further titles which are pertinent to this question can be found under the key-words "fluorescence-spectrum," "Europium-hexafluoroactylacetonate," "fluorescent," and "Europium-chelate." Many of the same titles found under "fluorescence," etc., are also found under the keyword "Europium." Do NOT record a document serial number more than once for any one question.

i. Continue your search for each question until you feel you have selected ALL pertinent documents to the question or have determined that no pertinent document can be found.

j. Record the time when you finish searching on this question.

As has been stated above, it may be necessary to look under various forms of the same word when locating information. For example, if one wants information on solubility, he can look under "solubility," "solubilities," "solubilization," and "soluble."

Following is a test question to be searched in order for you to familiarize yourself with the index. Please feel free to ask any question you might have.

Thank you for your cooperation in this test.

3. Sample Question for KWOC Index Test

Search in Index A:

Information on ion exchange in connection with research on water.

B. Index E

This index is a subject index, with citations listed under a subject heading assigned by the researcher.

At the front of the index, you will find a typewritten list of subject headings. The computer print-out is a list of subject heading

codes arranged in alphabetical order. All citations dealing with a particular subject are listed after each subject heading code.

In your search of this index you will:

1. Record the time starting search.

2. Read the question.

3. Look over the list of subject headings to determine under which ones you will search.

4. Record in the "keyword" column the subject heading code that you will search first.

5. Turn to that code in the index and read the citations listed.

6. Record the document serial numbers of citations that appear to be relevant to the question.

7. Repeat steps 4, 5, and 6 for any additional subject headings that might contain relevant documents.

8. Search until you feel ALL relevant documents have been located.

9. Record time of end of search.

If we use the sample search that was done for the instructions for Indexes A and B, "Need information on fluorescence of Europium," we can illustrate the search procedure. After reading the question, read over the list of subject headings. Heading "FV" (Lanthanides - Spectra) would probably contain some relevant material. Enter "FV" in the keyword column. Turn to "FV" in the index and read the citations of the documents listed. In the document serial number column, enter the 14 serial numbers of the relevant documents. On further reading of the list of subject headings, you decide that heading "BZ" (Complexes - Spectra) might also contain some relevant material. Enter "BZ" in the keyword column, turn to "BZ" in the index, read the document citations listed, and enter any

document serial numbers you feel are relevant. In this case, only one, "BZ-019," would be listed.

Please try this search for the sample question given in the instructions for Indexes A and B. If you have any questions, please ask.

DOCUMENT CONTROL DATA R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)

Florida State University
Library School
Tallahassee, Fla. 32306

2a. REPORT SECURITY CLASSIFICATION

2b. GROUP

3. REPORT TITLE

Tests of indexes. A comparison of keyword from title indexes with and without added keywords and a single access point per document alphabetic subject index

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

scientific; INTERIM

5. AUTHOR(S) (Last name, first name, initial)

Jahoda, G.
Stursa, Mary Lou

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
January 1969	55	9

8a. CONTRACT OR GRANT NO. AF-AFOSR-895-65	9a. ORIGINATOR'S REPORT NUMBER(S)
b. PROJECT NO. 9769-02	
c. 61445O1 F	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) AFOSR 69-0455TR
d. 681304	

10. AVAILABILITY/LIMITATION NOTICES

[redacted] unlimited.

11. SUPPLEMENTARY NOTES

Tech Other

12. SPONSORING MILITARY ACTIVITY

Air Force Office of Scientific Research
Directorate of Information Sciences
Arlington, Va. 22209

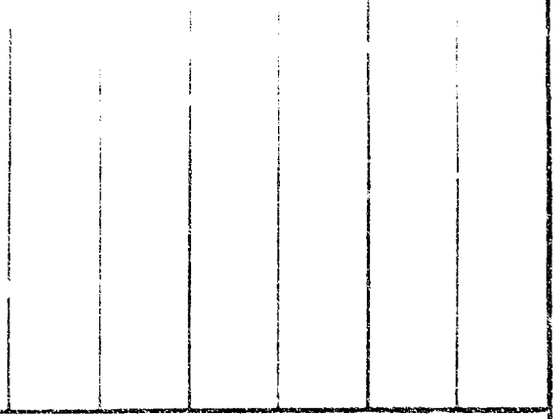
13. ABSTRACT

Three indexes to a collection of 3204 documents in the field of chemistry were test searched. The indexes are a keyword from title index without added keywords, a keyword from title index with added keywords, and a single access point per document alphabetic subject index. The three indexes were searched by 14 graduate chemistry students with 66 questions. Search results are characterized in terms of recall, precision, search time, and three other single figure measures. A measure of index performance based on recall, precision and search time was developed. There is no statistically significant difference between search results with the multiple access points per document keyword from title index and the single access point per document alphabetic subject index. Statistically significant differences in search results were found between the keyword from title index with added keywords and the keyword from title index without added keywords. The effect of the size of the delete word list and of elements of vocabulary control in keyword from title indexes were also studied.

DD FORM 1 JAN 64 1473

Security Classification

[redacted]
Relative term
Index depth
Index evaluation
Keyword from title index
Precision
Recall
Search time
Vocabulary control
INFORMATION STORAGE & RETRIEVAL
INFORMATION SYSTEMS



INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year, or month, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through _____."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through _____."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through _____."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13 ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS), (S), (C), or (U)*

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14 KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional.