A COMPARATIVE STUDY OF THE END USERS' AND THE INFORMATION EXPERTS' SEARCHING OF RLST'S DATABASES IN TERMS OF SEARCH TIME, RECALL AND PRECISION RATIOS

J. MEHRAD, Ph. D.
Dept. of Library and Information Sciences,
Shiraz University, Shiraz, I. R. of Iran
email: dean@srlst.com

A. SHOURE ZARI
Imam Khomeyni International University,
Qazvin, I. R. of Iran

Abstract - In this study, a comparison is made between the results of the end users' searching of RLST's databases and those of the information experts' searching in terms of search time, recall ratio and precision ratio. The study is primarily tasked with comparing the results retrieved through the three different search methods of direct, indirect and mixed. Five of RLST's information experts as well as RLST's direct users constituted the study's sample population. The research method and the data collection procedure used were semi-experimental method, observation and a questionnaire respectively. To test the hypotheses and to find an answer to the research questions, variance analysis, correlation coefficient as well as f, t and z tests were used at $\alpha = 0.05$.

Keywords - End user, Database, Regional Library of Science and Technology, Recall, Precision, Online Retrieval Service

INTRODUCTION

Since 1960s, computer technologies have increasingly been used worldwide to facilitate information processing. This use has already given birth to a number of new methods of information storage and retrieval. This process could be rightfully labeled as the starting point of a new kind of service known as 'On-line Retrieval Service', where users can browse the host's databases through a station connected to the network.

During 1980s, and particularly in developed countries, computer began to influence every aspect of the human life and consequently more and more people came to use this type of information service. During this decade, a number of databases came onto the market on CDs. Since these CDs had a storage capacity of 650 MB (equal to 200,000, and in more current types 300,000, A4 pages of information), database producers could successfully save the information of a whole database on such CDs. This storage method eliminated the telecommunication costs altogether, since it removed the need for being connected to the host's database.

Today, quite a few Iranian leading information centers like, *Iranian Information*

and Documentation Center, Jahad-e-Keshavarzi's Information Center and *Shiraz Regional Library of Science and Technology* produce and store a large number of their databases on CDs. In such cases, successful information retrieval depends on variables such as, choosing correct search strategy, the searcher's skill, knowledge and experience in searching and his acquaintance with the software in use.

In this research, carried out using the semi-experimental method, only some of RLST's databases, those retrievable using PC-SPIRS software, were considered. Having trained the end users of these databases, the researchers made a comparison between the results retrieved through the two search methods of direct and indirect.

PROBLEM STATEMENT

In developing countries, information is considered to be not only a national heritage but also a necessary element that plays an unignorable role in decision-making, transferring technology, pushing science and technology forward and in preventing researchers from doing mere duplications.

Today, a large number of information sources are purchasable on CDs. This fact is due to the establishment of a large number of computer databases and the wide spread use of CDs as excellent means of information storage. Being an important information institute in USA, DIALOG Information Services produced 300 databases on CDs before 1992 [1]. Nowadays, many important information networks are delivering their information productions in an electronic environment: 'Data Star' in Switzerland, 'Blaise', 'BLCMP' and 'Pergamon Infoline' in England and 'Libiris' in Sweden being just some examples [2]. This process in information delivery has urged the majority of database producers to train the end users and the intermediaries to search the databases directly and personally [3].

Concerning the effectiveness of such training courses, Mackibbon [4] states that a novice user may be able to search Medline, without the help of an information expert, only after attending a three-hour training course.

In this regard, Hariri says that in Iran, *Central Library of Iran's Medical Sciences University* has been one of the forerunners of providing direct search services within Medline. He writes, "This library's training courses for Medline began in 1993 and from the very first days, it was well received by students of medical sciences [5].

But RLST has already used the indirect method of information retrieval and has not trained its end users to search RLST's databases directly and personally.

In this study, the researchers first trained the end users of PC-SPIRS database, which has a structure different from that of Medline and covers almost all areas of the human knowledge. Then, their search results were compared with those of the information experts in terms of search time, recall and precision ratios.

RESEARCH HYPOTHESES

More than 50% of RLST's users are interested in attending a short-term training course

on direct search method.

By attending a short-term course on searching, the end users will be able to retrieve their information needs directly and personally.

RESEARCH QUESTIONS

1. Are there any significant differences between the quality of the documents retrieved by the information experts (indirect search) and the end users (direct search) in terms of search time, recall ratio and precision ratio?
2. If there are, how are they related to the following factors? skill in searching, subject knowledge, computer literacy and familiarity with the database language, i.e. English.
3. How far is the mixed method (searching that involves experts and the end users both) successful compared to the two other methods mentioned before?

METHODOLOGY

Having used the semi-experimental method, and through a cross sectional study, this research intended to compare the search results of the three different retrieval methods of direct, indirect and mixed.

The end users participated in practical and theoretical courses held from one to two sessions. When ready, the subjects were given a search form containing some search terms related to their majors and were asked to browse them in the database. The documents they retrieved proved that the training course had been successful.

Later, the end users' information needs were searched both by themselves and by the information experts through the search terms they had written down on RLST's search forms. After this stage, the results retrieved by the experts were compared with those retrieved by the end users in terms of recall ratio, precision ratio and search time.

Finally, a mixed search was also carried out. Here, only recall and precision criteria were taken into account. Search time was not considered since it was believed to be under the influence of the previous searches; that is, in this stage some of the relevant or irrelevant documents retrieved had already been retrieved by one or both groups and, thus, it was quite easy to establish a link between them.

SAMPLE POPULATION OF THE RESEARCH

In this research, two sample populations were used. The first sample population consisted of five of RLST's information experts. The second was composed of RLST's end users of Latin databases. Due to the large number of users that RLST had, only a representative group of them were selected using the following method.

Mixed sampling was used to select the participants of the study. To select the participants, the total number of RLST's clientele was first calculated for each season of the past two years separately. Since varying numbers of clientele had visited RLST in

different seasons, the ultimate number of participants included in each group was determined based on the total number of users in each season.

To select the participants, each week a different day was chosen. The users were, then, informed that they could retrieve their information needs independently and personally if they attended a two-session training course. From among 859 users, 53 were randomly selected. Three other users were further taken out on the belief that they couldn't represent their respective populations: one user was a high school student, another was a Ph. D student and still a third was an expert in the field of Library and Information Sciences in Mirza-ye Shirazi Library, Shiraz University, Iran.

From among these 50 users, 40 (from these 14, 6, 11 and 9 had visited RLST in Spring, Summer, Fall and Winter respectively) attended all the training courses.

The sample population under study is believed to be homogenous for two reasons: firstly, more than 95% of RLST's users belong to the educated class, or better, to the class of university students; secondly, users of Latin databases often master, or at least have some command of, the database language. These two factors made the researchers believe that no significant difference would occur in the results obtained if the number of participants was raised. Moreover, the number of participants was more than ten times as many as the number of the three variables of the study, namely search time, recall ratio and precision ratio. This reassured the external validity of the study.

STATISTICAL AND DATA ANALYSIS PROCEDURES

In this study, descriptive and inferential statistics (mean, standard deviation, variance, variance analysis, z-test, t-student test, f-test and correlation coefficient) were used to analyze the raw data using SPSS software.

DATA ANALYSIS AND RESEARCH FINDINGS
DATA ANALYSIS RELATED TO THE TRAINING COURSE

To assess the quality of the training course, a questionnaire was distributed among the trainees. The results, as shown in Table 1, indicate that 20% of the trainees evaluated the training course as fair, while 70% and 10% of them considered it to be good and very good respectively. No trainee evaluated the training course as weak or very weak.

Using $P_\circ \pm Z_{1-\frac{a}{2}} \sqrt{\frac{p_\circ q_\circ}{n}}$ it could be claimed that $80\% \pm 1.96 \ (0.063)$ of the trainees would evaluate the training course as good or very good if RLST introduced some qualitative courses to enable the users to handle the search process themselves. The results would be significant at $\alpha = 0.05$

In other words, between 79.88% and 80.12% of the trainees will feel content with the course and will assess it as good or very good.

Table 1: The frequency distribution of the trainees' evaluation of the course quality.

Course Quality	Frequency	Percentage	Cumulative Frecuency
Very Weak	-----	-----	-----
Weak	-----	-----	-----
Fair	8	20	20
Good	28	70	90
Very good	4	10	100
Total	40	100	

Table 2: The frequency distribution of the training course duration as shown by forty of RLST's end users (by minute)

Training Course Duration	Absolute Frequency	Absolute Percentage	Cumulative Frequency
30	1	2.5	2.5
40	5	12.5	15.0
45	3	7.5	22.5
50	4	10.0	32.5
60	2	5.0	37.5
90	1	2.5	40.0
110	2	5	45.0
120	3	7.5	52.5
150	1	2.5	55.0
160	1	2.5	57.5
180	3	7.5	65.0
185	1	2.5	67.5
190	1	2.5	70.0
200	3	7.5	77.5
215	1	2.5	80.0
220	4	10.0	90.0
240	4	2.0	100.0
Total	40	100	
Mean Course Duration: 130.875			Ucl = 155.33
Standard Deviation: 76.526			Lcl = 106.42
Standard Error: 12.100			

Table 2 shows that the time devoted to training varies considerably between 30 and 240 minutes. The resultant difference (210 minutes) is due to the different background knowledge that the trainees had about computer: those with little or no such background knowledge could personally handle the searching operation after 4 hours of training at

the outside.

Given $\overline{X}$ =130.875, SD=12.100 and using t distribution for finding confidence interval with α= 0.05, the mean course duration for RLST's direct users was calculated between 106.42 and 155.33 minutes using the following formula:

$$\overline{X} \pm t_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}} = 130.875 \pm (2.021)(12.100)$$

$$\nearrow \boxed{155}$$
$$\searrow \boxed{106}$$

Table 3 shows the degree of the participants' interest in extending the training course duration. Using the Likert Scale, each answer was scored between 1 and 5. The results indicate that the participants are not much interested in long-term training courses; rather, they prefer to learn the searching process through personal practice and repetition. In fact, they think that one or two sessions of training will suffice. Furthermore, they evaluated their own searches as very good, which was the highest choice possible in the very weak to the very good continuum.

Table 3: The degree of the participants' interest in extending the course duration

Questions Asked \ Rank	Very Little	A Little	Fair	Much	Very Much	Total	Mean
Interest in Extending the Course	18 45% 18	4 10% 8	10 25% 30	8 20% 32	0 0 0	40 100% 88	2.200
Learning Through Personal Practice and Repetition	0 0 0	9 22.5% 27	9 22.5% 52	13 32.5% 45	9 22.5% 45	40 100% 14.5	3.55
Evaluation of Personal Searches	0 0 0	0 0 0	8 20% 24	20 50% 80	12 30% 60	40 100% 164	4.100

Table 4: The frequency of the participants' search time after attending the training course.

Search Time	11	12	13	15	18	19	20	21	22	23	24	25	27	29	32	35	37	38	40	42	45	Total
Absolute Frequency	1	3	2	2	2	1	2	2	5	1	4	1	1	1	1	3	1	1	2	2	2	40
Absolute Percentage%	2.5	7.5	5.0	5.0	5.0	2.5	5.0	5.0	12.5	2.5	10.0	2.5	2.5	2.5	2.5	7.5	2.5	2.5	5.0	5.0	5.0	100
Cumulative Frequency%	2.5	10.0	15.0	20.0	25.0	27.5	32.5	37.5	50.0	52.5	62.5	65.0	67.5	70.0	72.5	80.0	82.5	85.5	90.0	95.0	100.0	100

$X_{UCL} = 179$ $X_{Lcl} = 22.179$ Variance = 101.631   Standard Error = 1.594   Standard Deviation = 10.081   Mean = 25.400

Since the mean and the standard deviation of the user population were unknown, William Seely Gosset's proposed formula was adopted to calculate the t distribution (with $\alpha = 0.05$ and $df = n-1 = 40 - 1 = 39$) as follows:

$$\bar{x} \pm t_{1-\frac{a}{2}} \frac{s}{\sqrt{n}} = 25.400 \pm (2.021)(1.594)$$

In other words, in 95% of the cases RLST users will be able to fulfil their own information needs within 22.179 and 28.621 minutes if they take part in similar training courses.

Table 5: The frequency of the retrieved documents and search duration as obtained by the five experts for 6 identical search queries

Experts	Frequency of Retrieved Documents	Search Time	Frequency of Retrieved Documents	Search Time	Frequency of Retrieved Documents	Search Time	Frequency of Retrieved Documents	Search Time	Frequency of Retrieved Documents	Search Time	Frequency of Retrieved Documents	Search Time	Total Retrieved Documents	Total Time	Mean Retrieved Documents	Mean Time	Standard Deviation of Retrieved Documents …	Standard Deviation of Search Time …
A	17	9	23	11	24	11	19	10	23	11	25	11	131	63	21.83	10.50	3.13	0.84
B	19	10	20	11	25	15	19	11	21	11	24	12	128	70	21.33	11.66	2.58	1.75
C	17	10	23	11	23	12	20	11	22	11	22	11	127	67	21.17	11.17	2.32	0.75
D	17	10	23	11	23	14	19	11	21	11	21	12	121	70	20.17	11.67	2.04	1.37
E	16	10	21	12	23	12	17	11	21	11	20	11	118	68	19.67	11.33	2.66	0.82
	$t^° = 1.05_{doc.}$								The Five Experts' Mean Retrieved Documents				The Five Experts' Mean Search Time		20.83	11.53		
	$t^° = 1.46_{tim}$								The Standard Deviation of the Five Experts' Retrieved Documents				Standard Deviation of Search Time…		0.79	0.71		

Considering the brightest searcher as the one who retrieves the highest number of documents in the least amount of time and the unskilled searcher as the one whose performance is quite the opposite, the researchers used directional hypothesis to show if there were any significant differences between/among the results obtained by different experts.

To use the test, the search results of the most skilled and the least skilled experts had to be compared. Finding no significant difference between the two extreme searchers will enable us to predict that there will be no difference between/among other searchers as well, since the search results of other experts will always fall somewhere between the search results of the two extreme searchers, the most skilled and the least skilled.

ANALYZING AND COMPARING THE SEARCH RESULTS RETRIEVED BY DIFFERENT INFORMATION EXPERTS

To be able to compare the search results obtained by the five information experts, it was necessary to use an appropriate statistical test. To do so, a single search query was randomly selected from each of the following fields: Sciences, Engineering, Agriculture, Veterinary Medicine, Medicine and Humanities. As a result, each expert was given six equal search queries. The results of the analysis are shown in Table 5.

As shown in Table 5, the number of documents retrieved by the most skilled expert has a maximum mean of $\bar{x}_{max}$= 21.83, with SD=3.13. The minimum mean obtained by the least skilled expert is $\bar{x}_{min}$=19.66, with SD=2.66.

To compare the above two means, one-tailed t-test (H_0: $\mu_1=\mu_2$ vs. H_1: $\mu_1 \rangle \mu_2$) was used with df=n1+n2-2 and α=0.05. The result, t=1.05, is smaller than t=1.860 and thus the null hypothesis cannot be nullified. That is, there are no significant differences among the means of the documents retrieved by RLST's experts.

The same test was also used to compare $\bar{x}$ max and $\bar{x}$ min for search time. The result, t*= 1.46, cannot reject the null hypothesis at α=0.05. That is, there is no significant difference between the two means obtained by RLST's experts for different subjects. Thus, it can be concluded that for comparing and analyzing the search results retrieved by the participants and by RLST's experts, it will not make any difference if one expert rather than another repeats the search operation all over again.

COMPARING THE EXPERTS' AND THE PARTICIPANTS' SEARCH RESULTS

Table 6: The frequency distribution of search time for 40 subjects as shown by the information experts

Search Time (by minute)	Frequency	Percentage	Cumulative Frequency
9	3	7.5	7.5
10	8	20.0	27.5
11	16	40.0	67.5
12	10	25.0	92.5
13	1	2.5	95.0
14	2	5.0	100.0
Total	40	100.0	
Mean: 11.100	Standard Deviation: 1.150	Standard Error: 0.182	Variance: 1.323

Table 6 illustrates the frequency distribution of search time as shown by RLST's information experts. Search time frequencies for the participants have already been discussed in Table 4. The results found in Tables 4 and 6 have been collapsed in Table 7

for ease of comparison and reference.

To test the differences between the mean search times as obtained by RLST's experts and the trained participants, S_2^2 and S_1^2 were used, since the variance and the mean of the population were not known.

Table 7: Mean, standard deviation and variance of the experts' and the participants' search times (by minute)

Variables \ Groups	Mean	Standard Deviation	Variance
Participants	25.400	10.081	101.631
Experts	11.100	1.150	1.323
	F_{ucl}=8.26	F*=76.19	F_{Lc}=0.319
		T*=8.538	T=1.684

F- test was used at α=0.05 to determine if there were any differences between the different variances. In this test, t*=8.538 and f*=76.19, were the two extreme critical values in the curve. Given f*=S_1^2/S_2^2=76.19, the hypothesis of no difference was rejected. That is, the difference between the unknown variances of the population was found to be significant. For this reason, the researchers made use of S_y^2 and S_x^2 in the t-test for the two independent samples with different variances since σ_x^2 and σ_x^2 were not only unknown but the difference between them was also significant.

Given the Satterth Waite's Method and at α=0.05, the researchers found out that t*=8.536, α=0.05, had occurred in the area where the null-hypothesis could be rejected. This means that we can be 95% sure that the participants' search time will be significantly longer than that of the experts.

One-tailed test for the difference between proportions in the two populations was used with $\dfrac{h_o : p_1 = p_2}{h_1 : p_1 > p_2}$ at α=0.05 to find out if there were any significant differences between the recall ratios of the documents retrieved by the experts and the participants.

To calculate Z*, $Z = \dfrac{\frac{k_1}{n_1} - \frac{k_2}{n_2}}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ was used since $n_1\hat{p}_1$, $n_1\hat{q}_1$ and $n_2\hat{p}_2$, $n_2\hat{q}_2$ were all larger than 5. Z*=2.5148 is larger than Z=1.64 and, thus, the null hypothesis is nullified. In other words, the recall ratio of the documents retrieved by the experts is larger than that retrieved by the participants.

This might be due to the fact that the information experts are not as informed about the information needs of the participants as the participants themselves. For this reason, they adopt search strategies that will possibly produce more results. Having received the retrieved documents, the participants will, then, be required to choose those that are more relevant.

Table 8: Comparing the recall ratios of the documents retrieved by the experts and the participants

Retrieved Documents Groups	Frequency	Overlap	Recall Ratio	
Experts	588	312	71.97%	$P_1=0.7197$ $Q_1=0.2803$
Participants	541	312	66.22	$P_2=0.6622$ $Q_2=0.3378$
Total	1129	Without Overlap817	Z=1.64	Z*=2.51

To see if there were any significant differences between the recall ratios of the documents retrieved by the two groups, one-tailed test for the differences between the proportions in the two populations was used with h_0: $p_1=p_2$ vs. h_1:$p_1 \rangle p_2$ at $\alpha=0.05$.

Here, the statistical variables used in the recall ratios difference test were utilized again. As shown in Table 9, Z*=2.17. This means that the difference found between the recall ratios of the documents retrieved by the two groups under study is significant.

Table 9. Comparing the precision ratios of the documents retrieved by the experts and the participants.

Retrieved Documents Groups	Frequency	Precision	Precision ratio	
Participants	588	449	82.99%	$P_1=0.8299$ $Q_1=0.1701$
Experts	541	458	77.89%	$P_2=0.7789$ $Q_2=0.2211$
Total	1129	907	Z=1.64	Z*=2.17

Table 9 shows that the precision ratio of the documents retrieved by the participants is more than that of the documents retrieved by the experts.

The participants' skill in subject area seems to have increased the precision ratio of their retrieved documents. In other words, the end users do their best to retrieve only those documents that are more related to the subject. In this regard, they are more successful than the information experts.

COMPARING MIXED, DIRECT AND INDIRECT METHODS OF INFORMATION RETRIEVAL

The investigations showed that the participants' retrieved documents had a precision ratio higher than that of the information experts.

In order to determine the differences between the precision ratio of the documents retrieved by the mixed method and that retrieved by the participants, Z-test was used at $\alpha=0.05$.

This test rejected the null-hypothesis. Thus, it can be said that there is a significant difference between the precision ratio of the documents retrieved by the participants and the mixed method. The precision ratio of the former was found to be higher.

The search results obtained by the participants, the experts and the mixed method have been summarized in Table 10.

Table 10: Comparing the search results retrieved by the three different methods of direct, indirect and mixed.

Retrieved Documents / Groups	Mean Recall Ratio	Mean Precision Ratio	Variance	Standard Deviation	Standard Error
Experts Z=1.64	14.700		14.421	3.797	0.600
	Z*=0.472	11.225	10.869	3.297	0.521
Participants		11.450	9.666	3.109	0.492
	13.525		16.820	4.101	0.648
Mixed		12.400	7.856	2.803	0.443
	13.875	Z*=3.140	9.169	3.028	0.479

Z-test was also applied, at $\alpha=0.05$, to compare the precision ratio of the documents retrieved by the two methods of mixed and indirect. The results showed no significant difference between the two ratios.

It can be concluded that mixed method produces the best results in information retrieval. In mixed method, the recall ratio of the documents retrieved is the highest, compared to the other methods. That is, it retrieves more documents. Moreover, the precision ratio of the retrieved documents is also high. This means that the least amount of false drops will be observed.

ANALYZING THE RELATIONSHIP BETWEEN THE RESEARCH DEPENDENT AND INDEPENDENT VARIABLES.

Table 11 shows the relationship between computer knowledge and database language literacy. The results indicate that the ratios related to database language literacy can

determine search time, recall ratio and precision ratio better than those related to computer knowledge. In other words, familiarity with the database language is more important than that with computer knowledge.

Typing speed, as an intervening variable, is a function of computer and database language literacy, which are prerequisites for successful search. Put it another way, skill in searching comes after computer and database language literacy.

Table 11. Comparing the ratios related to computer and database language literacy

	Searching	Recall Ratio	Precision Ratio
Database Language Literacy	0.6897	0.4264	0.4787
Computer Literacy	0.6534	0.3889	0.2650

Table 12 shows the relationship between search time and speed in typing. The significant level (0.0000) shows that search time and typing speed are strongly related. This means that those whose typing abilities are different will certainly (100%) retrieve the relevant documents in different times as well.

The coefficient of determination (95%) means that 95% of the differences found in search time are due to the participants' varying degrees of speed in typing.

Table 12. The relationship between typing speed and the retrieved documents' search time.

Degree of Freedom	Mean of Squares Variance	Sum of Squares	Change Source
24	158.2299	3797.5167	Between Groups
15	11.0722	166.0833	Within Each Group
39		3963.6000	Total
	Eta Squared= 0.9581	Eta=0.9788	
	F-Ratio=14.2907 Significant Level: 0.0000		

Tables 13 and 14 show the relationship between typing speed and the retrieved documents' recall and precision ratios. The significant level (0.1407) shows that the null-hypothesis cannot be rejected. That is, at $\alpha=0.05$ there is no significant difference between the mean precision ratios of the documents retrieved by those having different typing abilities.

Although the null-hypothesis has been rejected, we cannot conclude that there is a significant difference between those having different typing speeds and the mean recall ratios of the retrieved documents. In fact, some other factors might be involved here.

For example, those with varying degrees of computer knowledge will inevitably have different speeds in typing. Thus, a difference in computer knowledge may result in a difference in the recall ratio of the retrieved documents.

Table 13. The relationship between typing speed and the recall ratio of the retrieved documents

Degree of Freedom	Mean of Squares Variance	Sum of Squares	Change Sources
24	21.7767	522.6417	Between Groups
15	8.8889	133.3333	Within Groups
39	Eta Squared= 0. 7967	655.9750	Total
		Eta: 0.8926	
	F-ratio=2.4499 Significant Level: 0.0379		

Table 14. The relationship between typing speed and the precision ratio of the documents retrieved

Degree of Freedom	Mean of Squares Variance	Sum of Squares	Change Sources
24	11.5094	276.2250	Between Groups
15	6.7167	100.7500	Within Groups
39	Eta Squared= 0.7327	376.9750	Total
		Eta:0.8560	
	F-ratio=1.7136 Significant Level: 0.1407		

In order to analyze the relationship between other variables, their correlation coefficients are shown in the form of a matrix in Table 15 below.

Table 15 shows the correlation coefficient of the variables under study in matrix form. The correlation of these variables can be summarized as follows:

Computer knowledge is positively correlated with precision ratio and negatively with recall ratio, search time and typing speed.

Database language literacy is positively correlated with precision ratio and negatively with recall ratio, search time and typing speed. The positive correlation between language proficiency from one hand and recall and precision ratios on the other seems to be due to the fact that high language proficiency facilitates the application of search commands, promotes computer knowledge and subsequently changes the recall and precision ratios.

Table 15. The correlation coefficient of dependent and independent variables

Variables	Computer Knowledge	Language Proficiency Level	Precision Ratio	Recall Ratio	Search Time	Typing Speed
Computer Knowledge	1.000	0.4054	0.4257	-.5241	-.7771	-.8627
Language Proficiency Level	0.4054	1.000	0.6774	-.6373	-.8187	-.5795
Precision Ratio	0.4257	0.6774	1.000	-.8653	0.1554	-.5015
Recall Ratio	-0.5241	0.6373	-0.8653	1.000	0.4443	0.6449
Search Ratio	-0.7771	0.8187	-0.1554	0.4443	1.000	0.9000
Typing Speed	-0.8627	0.5795	-0.5015	0.6449	0.9000	1.000

CONCLUSIONS

Statistics showed that BA/BS and MA/MS students were the main users of RLST. Also, the sample population of this study primarily consisted of these students.

The results indicated that between 81% and 86% of the end users were interested in attending the training courses. This finding is in line with the first hypothesis of the study, which states, "the majority of RLST's end users are interested in participating in a short-term training course. All (100%) of the trainces evaluated the training course as fair or good. This course was evaluated as good or very good by 80% of the trainces.

The findings show that novice searchers can browse their information needs personally after attending, at most, four hours of training. Also, the mean of the training time varies between 110 and 156 minutes.

There were found certain differences between the documents retrieved by the end users and those retrieved by the experts. Some of these differences are mentioned below:

Search time for the information experts was significantly shorter than that for the end users. This is due to the fact that the end users are not as experienced as the experts in searching. It is believed that the end users can improve their search quality and reduce search time through practice and repetition.

The recall ratio of the documents retrieved by the information experts is higher than that by the end users. This is because the information experts, unlike the end users, are not well informed about the details of the subject to be searched. This urges the experts to make use of certain strategies to retrieve more documents. The retrieved documents

are, then, forwarded to the end users who have to choose the relevant documents. On the contrary, the precision ratio of the documents retrieved by the end users is higher than that of the documents retrieved by the experts. Familiarity with the subject area was believed to be responsible for such a difference.

Below, the variables of the study are shown in sequence, each preceding variable is more important than the next.

Database Language Literacy	→	Computer Knowledge	→	Scarching Skill	→	Familiarity with the Subject
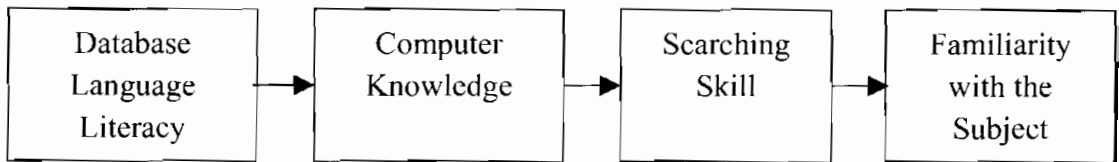
The results from the mixed method had the highest precision and recall ratios.

REFERENCES

[5] Hariri, N., *The results of direct and indirect search in Medlin,* Faslname-ye Ketab, No. 3, 1376/1997, pp. 26-37.

[2] Hunter, E., *Machine indexing,* Trans. by Mehrad, J. and Azarakhsh, M., Iranian Information and Documentation Center, Tehran, 1372/1994, pp. 185-210.

[1] Lambert, J. and Lambert, P. A., *How to get access to scientific and technical information?* Imam Hossein University, Tehran, 1372/1993, p. 62.

[3] Mall Li., Y., *The principles of information retrieval training,* Trans. by Sabokrooh, F. and Omrani, A., Vaje Pardaz, Tehran, 1376/1997, p. 11.

[4] Mckibbon, K., Ann et al., *How good are clinical Medline searches? a comparative study of clinical end-user and librarian searches.* Computer and Biomedical Research. No. 23, 1990, pp. 583-593.