MEASURES THAT DISCRIMINATE AMONG ONLINE SEARCHERS WITH DIFFERENT TRAINING AND EXPERIENCE

MESURES DISCRIMINANT LES CHERCHEURS AVEC DIFFIRENTES FORMATIONS ET EXPERIENCES EN TELEREFERENCE

Helen Howard
1602-55 Maitland Street
Toronto, Ontario M4Y 1C9

ABSTRACT

The major research objectives were to identify (1) the differences among the searches of users of online systems who have different backgrounds of training and experience, and (2) the differences between the searches of persons with and without experience on the ERIC data base.

Forty-two searchers separated into five experience levels performed two pre-selected searches on the DIALOG system. Variables that described the search process (e.g. number of commands used, number and types of errors) were measured by examination of the search transcripts.

RESUME

Les principaux objectifs de l'étude étaient d'identifier (1) les différences dans les recherches des analystes avec divers niveaux de formation et d'expérience en téléréférence, et (2) les différences dans les recherches des personnes avec et sans expérience préalable avec la banque Eric.

Quarante-deux chercheurs divisés en cinq niveaux d'expérience ont effectué chachun deux recherches pré-définies sur le système DIALOG. Les variables identifiant le processus de la recherche (e.g. le nombre de commandes utilisées, le nombre et le type d'erreurs) furent mesurées par l'analyse des transcriptions des recherches.

MEASURES THAT DISCRIMINATE

INTRODUCTION

The assumption exists that quality online search results - that is a bibliography that satisfies the end user - are not just products of chance. Quality searches and search results are thought to reflect the sum total of a searcher's efficiency and effectiveness. A second assumption is that the better trained and more experienced an online bibliographic searcher is, the higher the quality of the search.

OBJECTIVES OF RESEARCH

The major objectives of this research are to (1) identify the differences in search behaviour of users of online systems who have varying backgrounds of training and experience, and (2) the differences between the searches of persons with or without experience on the particular data base being searched.

Controlled research on the relationship between searchers' background of experience and training and the searches they perform is very limited, (for example see Fenichel, 1981; Howard, 1980; Wanger, 1980). Much discussion has taken place on the question of what makes a "good" searcher. Traits and skills which have been debated include personality and mental qualifications; subject knowledge; knowledge and experience with reference tools, particular online systems, and data bases; and spelling and typing skills. Central to these issues are the problems of identifying appropriate training and experience.

METHOD

A quasi-experimental design was developed with forty-two searchers from five groups: novice searchers, moderately experienced searchers without ERIC data base experience, moderately experienced searchers with ERIC experience, very experienced searchers without ERIC experience, and very experienced searchers with ERIC experience. The experience level was determined by three criteria: the length of time each subject had been searching, the total number of searches ever conducted, and the average number of searches per month over the last six-month period. It is important to note that the definition of a search in this study is one used frequently but not exclusively, i.e., the accessing of one data base in response to a search inquiry.

The searchers worked in twenty-two different institutions. Their distribution by type of institution is shown in Table I. Data on searcher background such as education, training, experience, as well as institutional policies regarding charging were collected during individual interviews. Very explicit written instructions on the experimental procedures were given to

the subjects after the interview. They then had an opportunity to raise questions to clarify the instructions. Two pre-determined search requests in narrative form along with a short questionnaire requesting information on such items as whether the search inquiry statement was adequate, whether any aspect of the search was difficult, and if there had been any mechanical problems were left in a sealed envelope with each subject.

TABLE I

DISTRIBUTION OF SUBJECTS BY EXPERIENCE AND ORGANIZATION

Organization	Novice		Moderately Experienced No-ERIC		Experienced ERIC		Very Experienced No-ERIC		Experienced ERIC		Total	
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
University	7	(16.7)	4	(7.1)	5	(11.9)	2	(4.8)	5	(11.9)	22	(52.4)
Special Library	1	(2.4)	3	(7.1)	1	(2.4)	3	(7.1)	0	(0)	8	(19.1)
Government	4	(9.5)	1	(2.4)	1	(2.4)	1	(2.4)	2	(4.8)	9	(21.4)
Other	2	(4.8)	0	(0)	0	(0)	1	(2.4)	0	(0)	3	(7.1)

The search instructions included the typing of five ED's and five EJ's online in format three and printing up to fifty of each offline in format five. The subjects mailed their printouts (called transcripts) of their searches to me and Lockheed sent me the offline printouts. The transcripts and printouts provided the raw material for analyzing searcher behaviour and for determining precision and recall.

The experiment was designed to control as many of the environmental variables as possible: the system (DIALOG); the data base (ERIC); search topic; type of terminal; availability of the ERIC thesaurus; system response time (searching to be done in off-peak hours); length of bibliography; total cost of search; and the absence of inquirer throughout the search process.

The two search inquiries were developed from topics on which I had been doing research: (1) the collective bargaining of librarians; and (2) innovations in libraries since 1973. It is difficult to determine criteria for judging the level of difficulty of a search. I treated the two search inquiries as bona fide ones which I would have made. In my view Search A was precise and limited in scope and therefore relatively easy to plan and execute. The narrative of Search B contained more concepts and relationships among the concepts. Consequently I judged it to be moderately difficult.

Some fifty variables, called search process variables, were created for analyzing the searches. The variables were divided into three categories to measure: (1) efficiency in using the DIALOG system; (2) efficiency in using the ERIC data base; and (3) error rate. The last category was subdivided into clerical errors such as misspelling, and others such as errors in the command language, procedural errors, and errors in logic.

Outcome variables of cost effectiveness, precision, and recall were also calculated (Howard, 1980). An estimated recall figure had to be used because there was no way of knowing how many relevant references were in the data base. The estimate was arrived at by keeping a record of all items judged to be relevant and then totalling them.

RESULTS

The data are being analyzed to provide descriptive statistics, cross-tabulations, and analyses of variance. When this paper was written, the statistical analyses were still incomplete. Additional results as well as fuller interpretation will be reported at the CAIS Conference.

SEARCH PROCESS VARIABLES

MEASURES THAT DISCRIMINATE

The mean values for selected search process variables are presented in Tables II and III. The differences between the results of the two searches are striking. No clear-cut pattern of progression across the groups emerges. In Search B there is the greatest consistency of performance with the Very Experienced, ERIC searchers displaying the largest number of sets, using the most commands, the largest number of thesaurus terms and of free text terms. This suggests that as a group they exerted the most effort in the search process. In general, the Moderately Experienced, ERIC group has noticeably lower values for these variables. The Moderately Experienced, No-ERIC group tends to have still lower values for these variables and the Novices fall into a middle range.

TABLE II

MEAN VALUES FOR SELECTED SEARCH PROCESS VARIABLES: SEARCH A

Variable	Novice (N=14)	Moderately Experienced No-ERIC (N=7)	Moderately Experienced ERIC (N=7)	Very Experienced No-ERIC (N=7)	Very Experienced ERIC (N=7)
Commands used	10.86	20.29	11.29	15.57	16.50
Select	5.07	14.71	6.00	7.43	9.14
Limit	2.07	2.14	2.29	2.29	2.86
Combine	3.14	3.43	2.86	5.71	5.00
Expand	0.57	0.00	0.14	0.14	0.00
Sets viewed	10.50	19.86	10.86	15.86	17.38
Descriptors searched					
Thesaurus terms	3.29	2.43	5.00	6.29	7.43
Free text terms	1.86	12.29	1.29	2.29	2.29
Clerical errors	1.36	1.43	1.14	1.86	1.17
Connect time (minutes)	9.81	12.49	7.27	10.70	9.27
Preparation time (minutes)	19.64	21.14	22.14	28.57	13.57
Speed	1.16	1.38	1.59	1.56	1.83
System efficiency score	2.64	2.57	3.29	2.14	1.83
Data base efficiency score	2.14	2.00	2.00	1.14	2.00
Errors, dubious practice score	2.50	2.14	1.57	2.14	1.67
Total efficiency/errors score	7.29	6.71	6.86	5.43	5.50

TABLE III

MEAN VALUES FOR SELECTED SEARCH PROCESS VARIABLES: SEARCH B

Variable	Novice (N=14)	Moderately Experienced		Very Experienced	
		No-ERIC (N=7)	ERIC (N-7)	No-ERIC (N=7)	ERIC (N=7)
Commands used	22.75	15.86	19.67	27.86	25.50
Select	12.64	11.14	12.00	14.43	18.29
Limit	2.57	2.00	2.86	3.14	3.57
Combine	6.86	2.43	6.43	9.71	6.71
Expand	1.17	0.29	0.17	0.57	0.50
Sets viewed	21.93	15.57	21.57	27.86	28.86
Descriptors searched					
Thesaurus terms	8.64	9.00	11.86	17.00	12.29
Free text terms	4.64	2.57	1.29	2.57	6.29
Clerical errors	2.08	1.14	2.14	2.58	1.50
Connect time (minutes)	19.91	11.83	11.16	20.91	12.40
Preparation time (minutes)	43.21	30.00	30.71	67.14	30.29
Speed	1.21	1.42	1.83	1.39	1.98
System efficiency score	2.93	3.29	2.71	2.42	2.67
Data base efficiency score	1.36	1.14	0.86	1.00	0.86
Errors, dubious practice score	2.08	2.00	2.00	1.57	1.86
Total efficiency/errors score	6.31	6.43	5.57	5.00	5.50

In Search A there is considerable variability both across and within experience groups. For example, the Moderately Experienced, No-ERIC group which displays the largest number of sets and uses the most commands, also uses the fewest thesaurus terms and the largest number of free text terms. The reason why the No-ERIC group uses the most free text terms in Search B as compared with the most thesaurus terms in Search A is not clear. Variables related to the searches themselves may be affecting searching behaviour.

Although the mean number of clerical errors seems low with a range of 1.14 to 1.86 in Search A and 1.4 to 2.58 in Search B, 66 and 70 percent of the searchers made clerical errors in the respective searches. An unexpected finding was that the Very Experienced, No-ERIC group rather than Novices make the most clerical errors in both searches.

In respect to connect time in both searches, the Moderately Experienced, ERIC searchers spent the least time online and the Very Experienced, No-ERIC searches the most time. It is worth noting that mean connect times and mean preparation times are noticeably higher for Search B, the more difficult search.

Search Outcome Variables

All existing methods for measuring the quality of online searches are flawed. The three measures - cost effectiveness, precision, and recall - were chosen for this study because they are the best that are available. The mean values for the search outcome variables are presented in Tables IV and V.

In Search A and Search B the Very Experienced, ERIC group produced the most cost effective searches both when measured by dollar cost per relevant reference retrieved and by minutes of searcher time. The results for precision and recall are mixed across the groups. In Search A the Novices achieved the highest precision ratio while the Very Experienced, No-ERIC searchers achieved the lowest. The latter group, however, show the highest recall ratio and the Moderately Experienced, No-ERIC searchers the lowest recall ratio. One might suggest that it is reasonable for a searcher unfamiliar with the ERIC data base to conceptualize a search to produce a high recall rate but this does not explain the low rate for the Moderately Experienced No-ERIC group.

Search B results reveal somewhat more of a pattern with the Very Experienced, ERIC group producing the most cost-effective searches as well as the highest precision and recall ratios.

TABLE IV

MEAN VALUES FOR OUTCOME VARIABLES: SEARCH A

Variable	Novice (N=14)	Moderately Experienced No-ERIC (N=7)	Moderately Experienced ERIC (N=7)	Very Experienced No-ERIC (N=7)	Very Experienced ERIC (N=7)
Unit cost in $	.44	.67	.73	.41	.36
Unit cost in searcher minutes	19.64	21.14	22.14	28.57	13.57
No. relevant references retrieved	38.36	32.43	33.29	49.57	38.43
Precision	94.41	80.82	81.00	83.22	82.86
Recall	38.74	32.76	33.62	50.07	38.82

Estimated relevant references in data base: 99.0

TABLE V

MEAN VALUES FOR OUTCOME VARIABLES: SEARCH B

Variable	Novice (N=14)	Moderately Experienced No-ERIC (N=7)	Moderately Experienced ERIC (N=7)	Very Experienced No-ERIC (N=7)	Very Experienced ERIC (N=7)
Unit cost in $	1.85	1.14	1.00	1.61	0.88
Unit cost in searcher minutes	4.47	2.81	2.40	4.59	1.77
No. relevant references retrieved	22.79	24.29	21.43	24.00	27.57
Precision	27.79	35.23	47.57	27.64	51.61
Recall	9.23	9.83	8.68	9.72	11.16

Estimated relevant references in data base: 247.0

MEASURES THAT DISCRIMINATE

One-way and two-way analysis of variance tests were carried out on combinations of separate searches, combined searches, by ERIC group, and by experience group. At the time of this writing two significant relationships have shown up in the two-way ANOVA's: (1) In Search B there is a statistically significant relationship (at the 0.033 level) between precision and ERIC experience; (2) when the two searches are combined there is a statistically significant relationship (at the 0.024 level) between recall and ERIC experience

TRAINING

Of the forty-two subjects only six did not have formal DIALOG introductory training. The interview schedule recorded thirteen types of training. In an earlier data analysis no relationship could be shown between the kind of introductory training and the outcome variables. The results regarding advanced or up-date training are inconclusive. At this state of the analysis the type of training does not seem to be related significantly to either search process variables or outcome variables. Yet approximately thirty percent of the subjects indicated that they would like more training and most preferred supervised practice as a training method.

CONCLUSION

The preliminary findings of this exploratory study have several implications for groups providing online searching, training and/or developing training materials and for employers of data base searchers. For example, about forty-five percent of the subjects wanted more training, but not of the same type they had had. They preferred practice sessions which would be supervised and which would permit exploring different strategies in conducting a search.

Another implication is that a great many searchers lack opportunities for adequate practice. Nearly forty-five and sixty-five percent of the subjects felt that they did not get enough practice on the DIALOG system and the ERIC data base respectively to achieve and/or retain proficiency. The fact that browsing online was almost non-existent in the searches is an implication for the need to stress in training sessions the efficacy of this practice.

In general these preliminary results indicate that searching is a highly individualized process. They point up the particular need for more work to be done on the interrelationships of search process and outcome variables and how various search strategies will produce different results to meet specific needs.

REFERENCES

Fenichel, Carol Hanson. "Online Searching: Measures that Discriminate Among Users with Different Types of Experiences," in Journal of the American Society for Information Science, vol. 32 (January 1981), pp. 23-32

Howard, Helen. "Experience and Training for Online Bibliographic Searchers," in Proceedings of the Eighth Annual Canadian Conference on Information Science, Toronto, Ontario, May 6 - 10, 1980. 18 p. (microfiche)

Wanger, Judith. "Evaluation of the Online Search Process," presented at the National Online Conference, New York City, March 26, 1980. 8 p.