

STUDENT USE OF MEDLINE: AN ANALYSIS OF THE EFFECTS OF EXPERIENCE AND SEARCHING KNOWLEDGE AND SKILLS ON RETRIEVAL IN A CLINICAL SITUATION

Miranda Lee Pao, School of Information and Library Studies
Suzanne F. Grefsheim, Mel L. Barclay, Barbara L. Shipman
University of Michigan Medical Center

University of Michigan
Ann Arbor, Michigan

As a result of the rapid growth of new knowledge in the biomedical sciences, the Panel on the General Professional Education of the Physician (GPEP) recommended that medical schools train students to be life long learners. An important component of life long learning for physicians is the ability to identify and critically appraise the biomedical literature. It has been found that medical students, as other adult learners, learn these skills best when instruction is viewed as aiding in the resolution of a real information need. For this reason, the most effective inducement for students to learn to search and to use MEDLINE in their clinical decision making is when search assignments are made part of clinical rotations. (1) Further, it is assumed that reducing or eliminating barriers such as cost and lack of availability and providing a user friendly search interface make literature searching more accessible for students.

The University of Michigan makes the full MEDLINE file available at no cost to all students, faculty and staff 24 hours a day from any computer which can connect to the campus network via secondary communication processor, ethernet, or modem. Locally known as UM-MEDLINE, the search interface is actually PaperChase, which has been judged one of the most easy to learn end user systems for accessing MEDLINE available.(2) Students at the University of Michigan are given the opportunity to attend a UM-MEDLINE training session in their second year, but few take the opportunity to refresh or update their skills prior to their third year clinical rotations when they are likely to receive search assignments. This situation led to the following study.

RESEARCH DESIGN

A group of 49 third year medical students in an OB/GYN rotation was selected for the study. As part of an examination, they were presented with the following case:

You are a physician working in the emergency room in a small northern Michigan community. Mrs. HGW, a 39 year old nulligravid woman is vacationing in your area and comes to the emergency room complaining of severe abdominal pain. Upon taking a history you find that the patient has had ovulation induction with gonadotropins at the University within the last four weeks. On physical

examination you find that she has marked ascites and that when she is supine she is somewhat hypotensive. She reports about 20 lbs of weight gain in the last five days. What treatment modalities have been recommended in these circumstances?

The students were asked to use UM-MEDLINE "to retrieve those abstracts which would be most salient to the management of the patient's problem". Each individual was required to print the abstracts of the most relevant items retrieved from their search. The objective was to retrieve a few items which contained the needed information to manage the clinical emergency presented in the vignette. Thus a high precision subject search was called for. In addition to the printouts submitted by the students, PaperChase provided us with transaction logs for the searches. These have also been used as sample data. Although content of the transaction logs cannot reveal user motivation and satisfaction, the transcripts capture a number of useful objective data.(3,4)

Studies have shown that, intuitively, searching experience influences searching quality. In other words, if a physician has somehow developed a habit of consulting the UM-MEDLINE system, MEDLINE searching is more likely to be an option when there is a need for information.[4, 5] Furthermore, King has found that a "success-breeds-success" mechanism may be at work.(6) And Horowitz postulated that users tend to use more advanced search features as more experience on PaperChase is accumulated.(7) From these studies one would conclude that as an end user gains more search experience, s/he would become a more skillful searcher. Skill could take several forms. One could become more adapt in the use of a system's features so that less time was needed to execute searches. Fewer errors might be made. More of a system's advanced search features could be used in the formulation of search strategies. If they used appropriate Boolean operators and subject headings, searchers might conduct more high recall or high precision searches. Most features in PaperChase are designed to promote ease of use, and to complement and lead the casual user to more appropriate terms. Others inherent to the MEDLINE file are meant to focus on better description of document content so that more precise and more relevant papers may be retrieved.

This study is designed to address the following questions:

A. Does past use of the system relate to subsequent use of UM-MEDLINE? In other words, how likely is it that students required to search UM-MEDLINE during their clinical rotation will make heavy use, as opposed to low or no use of UM-MEDLINE, if they had been low or non-users previously?

B. When "critical" information is needed to manage patient problems, how effective are the searches performed by medical students?

The following variables were examined in this study.

Search Experience King found that the number of searches conducted rather than the length of time a user has searched is a better measure of a searcher's experience, at least in the use of UM-MEDLINE. (6) In this study, search experience is defined as the number of online sessions undertaken by an individual student at the time of the examination. Although the number of searches conducted might be a better indicator, it is possible for two or more searches to have been conducted on a single PaperChase session. A quick review of the sessions in hand, however, indicated that the students generally conducted only one search per session. As was done in King's study, three levels of experience were examined.

The study seeks to answer the question: does past use of the system relate to subsequent use? Therefore, the level of usage of UM-MEDLINE after the examination was measured by noting the number of search sessions conducted during the five months after the examination. System use was then compared among the users found in the three levels of search experience.

Search Effectiveness Search effectiveness calls for a certain level of searching knowledge and skill. At the basic level, judicious use of such search features as the Boolean AND and OR, or use of the PaperChase SCAN and DISPLAY options would tend to improve search effectiveness. In addition, UM-MEDLINE (PaperChase) contains a number of advanced search features which are not presented as initial menu choices. For example, the availability of the Boolean NOT as an option is only indicated in the users manual, to which most UM-MEDLINE users do not have access. The ability to use some limiter features, subheadings, and related subject headings is also not immediately apparent. Use of such search features should give an indication of the level of search skills, and the degree of sophistication a user possesses in the use of the system. Furthermore, although the use of MeSH headings is usually preferred, the appropriate use of title words can be effective under certain situations. An effective searcher appropriately applies his or her knowledge of the topic, the structure of the data file, and the search system. For example, if a

knowledgeable user retrieved a single hit using the term "ASCITETREATMENT", s/he would presumably assume that this is a typographical error in the title and that items on this topic must be posted under another term. Content analysis of the search transcripts was performed to identify whether the student searchers had this level of knowledge and skill and to reveal any missed opportunities.

Retrieval Effectiveness Retrieval effectiveness directly reflects on the quality of the search results. It addresses how well the search is able to retrieve items which could be used to solve the clinical problem. Two common performance measures, recall and precision ratios, were used. Computation of precision is straight forward. However, recall ratios require the knowledge of ALL relevant items contained in the database on the requested topic. This is impossible. Therefore relative recall was used. Two comparisons were made: retrieval from each student was compared with (1) a "quick and dirty" search from a subject expert searcher; and (2) the combined search results from both the subject expert and a search expert. The latter is similar to the procedure used by Haynes (8). Items in the two parallel retrieved sets were evaluated for relevance by the subject expert. Eliminating the duplicates, the relevant items identified from the pooled set were used to assess the retrieval effectiveness of the student searches.

RESULTS

Frequency of System Use The 49 medical students averaged 46.1 search sessions with a median of 40 at the time of the search assignment. Since UM-MEDLINE was introduced in January 1989, at the time of the examination it had been available for 30 months. During that period, therefore, these students logged only an average of 1.53 session per month. However, the overall usage of the system varied widely among different members of the group. Three students conducted their very first search at the time of the examination; 9 students had conducted 80 - 170 searches. Overall, the ready availability of UM-MEDLINE during the first 30 months had produced light use of the database. This bears out the findings of several other studies. [6,9,10,11] However, even this low average use of MEDLINE by students is higher than was reported by King in her study at the same institution after only six months of system availability. (6) The increase in average use by students could indicate that the pervasive presence of the system encourages use over time.

To determine the effect of experience level on subsequent UM-MEDLINE usage, the number of search sessions performed by the medical students in the 5 months following the examination was collected. The students had been placed into one of three experience groups based on the number of searches they had performed at the time of the examination: (1) beginners - those who had performed 1-20 sessions (n=17); (2) intermediates - those

who had performed 21-40 sessions (n=8); and (3) advanced - those who had performed 41 or more sessions (n=24). The choice of these ranges was based on the median number of search sessions.

At the time of the examination, students in the beginners group had conducted an average of 10 search sessions during the previous 30 months, whereas the intermediates and advanced searchers averaged 29 and 76 sessions respectively. Table 1 shows that although the beginners and intermediates averaged only one-third and one search session per month respectively, the advanced searchers averaged 2.5 sessions. An ANOVA test shows that a

Table 1: Average Search Sessions per Searcher per Month Before and After the Search Assignment Among Three Experience Levels

| | Jan 89-July 91 | July 91-Nov 91 |
|--------------|----------------|----------------|
| Beginner | 0.34 | 1.38 |
| Intermediate | 0.96 | 2.45 |
| Advanced | 2.52 | 5.02 |

statistically significant difference among the mean number of sessions for the three groups remains after 5 months. Indeed, the average number of sessions for the 5 month period is 6.88 for beginners, 12.25 for intermediate level searchers and more than twice that, 25.08, for advanced level student searchers. Moreover, a correlation coefficient of 0.75 was computed for the number of search sessions at the time of the search assignment and those conducted in the five months following the intervention for each student. Therefore, there is a strong indication that previous experience relates to greater use of UM-MEDLINE. However, while heavy users almost doubled

Table 2: Average Use of Search Features

| | King | Pao, et al. |
|----------------------|-------------------|--------------------------|
| Search Statements | 9 | 12 ± 6.3 (2 - 32) |
| Displayed or Scanned | 36.3 (0 - 449) | 32.8 ± 40.3 (1 - 227) |
| Printed | - | 7.4 ± 3.6 (0 - 15) |
| “AND” | 3.5 ± 3.9 | 4.1 ± 3.5 (0 - 14) |
| “OR” | 0.6 ± 1.1 | 1.0 ± 1.2 (0 - 5) |

Note: The values given in each cell are mean, standard deviation and range respectively.

their monthly searching levels, intermediate level searchers attained searching levels equal to those placed in the heavy use category prior to the examination, and beginners increased their searching by a factor of four.

Use of System Features The use of search features on the UM-MEDLINE system during the first year was reported in an earlier study of a sample of 50 medical students.[6] Table 2 and Figure 1 show comparative data from that sample with the present class of 49 third year students. In the present study there are, on average, 12 search statements with a median of 11, for this topic. This is one-third more than King found. Similar differences were found in the use of the Boolean AND and OR. Also, all students used or were directed to use medical subject headings. If the use of the controlled vocabulary is desirable, these results provide confirming evidence that PaperChase is able to promote use of MeSH terms. Since a “quick and dirty” search was expected, it was surprising to find that only 57% of this group used title words as compared with 90% two years ago. King reported the use of the limiting features such as

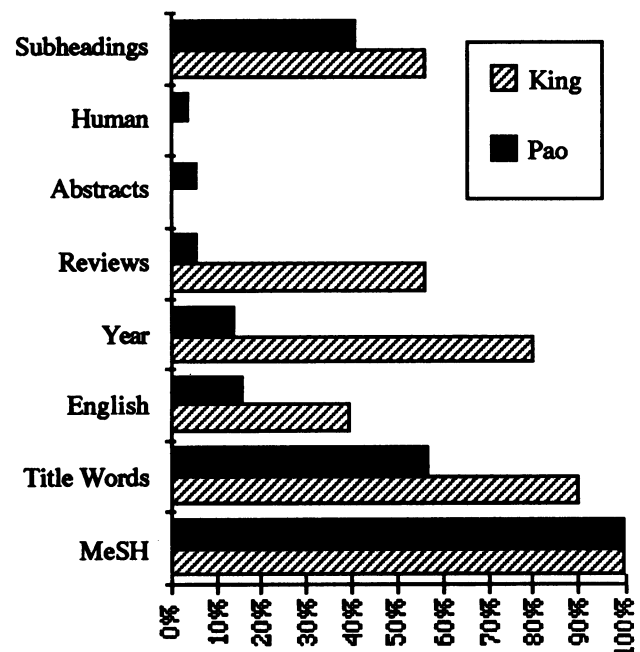


Figure 1. Percentage of Searchers Using Each of the Search Features

ENGLISH, YEAR, REVIEWS, and ABSTRACTS using both the individual search and the searcher as units of analysis. [6] In our study, only searchers were considered since only a single topic was examined. King reported much higher levels of use of limiting features. For example, the limiters ENGLISH, YEAR and REVIEWS were used by only a few of the searchers in our study, as compared to 40%, 80% and 56% respectively by the

searchers in King's study. While 41% of our students did use at least one subheading, King reported that 56% of the searchers in her study used subheadings. This difference may be due in part to the nature of the search topic under consideration.

A series of one-way ANOVA analyses were performed to examine if there was any significant difference in the use of each of the limiting features among the 3 experience groups. No statistically significant differences were found. However, if beginners were defined as those who conducted 1-10 sessions at the time of the examination, a statistical difference was found for the use of subheadings at the 0.05 level of significance ($F=4.04, p=0.024$). In other words, significantly fewer subheadings were used at the very early stage of UM-MEDLINE use, however, use of limiters overall is infrequent.

Retrieval Effectiveness In order to ascertain the retrieval effectiveness of the student searchers, an experienced online searcher was asked to conduct a high recall search on the same topic. Her search retrieved 14 papers compared to 7 citations retrieved by the subject expert's quick search. Only two papers were retrieved by both searchers. As a result a total of 19 unique items was retrieved. These unique items were rated by the subject expert for relevance. An "A" was assigned to definitely relevant items and a "B" to partially relevant papers. Ten were rated relevant of which 9 were definitely relevant. Five of the 7 citations retrieved by the subject expert were judged definitely relevant as were 6 of the 14 retrieved by the search expert. One additional paper retrieved by the search expert was rated partially relevant. Using the pooled relevant retrieval ($n=10$) as the basis for

Table 3: Relative Recall and Precision

| | Pooled relevant set n = 10 | | Definitely relevant set n = 9 | |
|----------------|-------------------------------|-----------|----------------------------------|-----------|
| | recall | precision | recall | precision |
| Subject expert | 50% | 71% | 56% | 71% |
| Search expert | 70% | 50% | 67% | 43% |
| Overlap | 20% | 100% | 22% | 100% |

comparison, the subject expert and search expert achieved relative recalls of 50% and 70% respectively and precision ratios of 71% and 50% respectively. (Table 3) While the subject expert achieved greater precision, the librarian retrieved more relevant articles.

By contrast, the relative recall from the students was 22%, with a median of 20%, when compared with the subject expert's quick search. (Table 4) The average improved to

27%, with a median of 30%, when compared with the pooled set of relevant items as well as when compared to the set of items rated of definite relevance. The students' retrieved sets ranged from zero to 13 articles with a mean of 7.4. Their precision ratio compared to the search expert was 17%, with one out of the seven citations being relevant to the question. That one article, however, had a good chance of being highly relevant. If one looks at the pooled set of relevant articles, the students' precision improves to 42% with about 3 of the seven articles rated relevant.

Table 4: Relative Recall and Precision by Medical Students

| | Expert retrieval n = 5 | | Pooled retrieval n = 10 | | Retrieval of definite relevance n = 9 | |
|-----------|---------------------------|--------|----------------------------|--------|--|--------|
| | mean | median | mean | median | mean | median |
| Recall | 22% | 20% | 27% | 30% | 27% | 22% |
| Precision | 17% | 18% | 42% | 50% | 39% | 41% |

*Based on 49 student searchers

Search Effectiveness In order to examine the concept of search effectiveness, it is necessary to understand why there was such variance among the retrieved sets produced by the subject expert and the search expert. One reason is that the subject expert used the term "menotropins" in his search as a result of a PaperChase prompt after entering "gonadotropins", the word used in the patient scenario. "Menotropins" is a more specific drug term than "gonadotropins". No one else used this term. As a result, a relevant paper published in 1974 was missed by every other search. Another relevant paper published in 1975 which the subject expert retrieved was retrieved by only 2 of the students. Consequently, using the subject expert's retrieved set as the basis for comparison, the students' recall ratios tended to be lower.

On the other hand, in conducting a high recall search, the search expert retrieved the most recent relevant paper published at the time. It was not retrieved by the subject expert. Curiously, of the students who retrieved at least one relevant paper, only 2 of them did not retrieve this one.

Most students were able to retrieve a few useful citations. Nearly 43% of them ($n=21$) were able to identify "ascites" and "ovulation induction" as two of the main concepts involved in the question and then proceed to combine the two headings retrieving nine citations. Five of the nine papers retrieved using this strategy were judged relevant. However, it should be noted that two of these relevant articles were not retrieved by the subject expert but they were retrieved by the search expert.

A less effective approach to the topic employed a strategy which combined the two subject terms "ascites" and "gonadotropins". Over 100 citations resulted. Several students proceeded to scan the list and select those they thought were relevant. Typographical errors in the database were spotted by a number of students. For example, several recognized "ascitetreatment" as a typo even though there was a posting under it and they proceeded to combine by ORing this title word with the results from the correct MeSH heading.

Thirty-nine, or 80% of all the students in the study, did retrieve at least one relevant citation. The fact that 10 students retrieved no relevant citations at all is troublesome. Of these 10, seven were in the beginner category, one was in the intermediate category and 2 were advanced users. Thus, 41% of the beginners retrieved nothing; as did 12% of the intermediate level searchers and 8% of the advanced group. An examination of the transaction logs of these failed searches showed a few trends. First, this search assignment was the first UM-MEDLINE experience for three of the people who did not retrieve any relevant citations. Yet, mere lack of search experience did not account for the failure of the rest.

The most consistent indicator of a failed search was the apparent lack of a logical search plan. It appears that these searches were based on terms rather than the concepts identified in the patient problem scenario. As a result, the searchers tended to begin by ANDing the three main subject terms, and, when none or only one "hit" was retrieved from such a strategy, they were unable to expand or modify the strategy to get at the basic concepts. An experienced searcher knows that each term should be considered merely one of several ways to represent a concept. These students under-utilized the Boolean OR, which could bring together all terms related to a concept, before using the Boolean AND to combine and limit the fully identified concepts. As a result many opportunities were missed.

There were also indications that subheadings and the use of limiters were not well understood. As an example, a subject heading was ANDed together with one of its own subsets, that is, the same heading was limited by a subheading. And limiters such as ENGLISH or a YEAR were often applied too early in the search. Finally, when a large set was retrieved, processing slowed considerably, particularly when the system was close to capacity. It appears that when this occurred, some searchers did not understand what was happening and they gave up their original search plan, at which point they floundered.

CONCLUSION

In conclusion, to answer the questions posed by the study, while the probability that heavy users would continue to be heavy users as compared to low or moderate users after the examination search experience is high, all categories

of user increased their use of UM-MEDLINE in the 5 months following the examination, with beginners increasing their use by a factor of four. Finally, in answer to the question whether the searches performed by students are effective, it can be said that 80% did retrieve at least one relevant citation on which they could have made a clinical decision in an emergency.

References

1. Frasca MA et al. A multidisciplinary approach to information management and critical appraisal instruction: a controlled study. *Bull Med Libr Assoc* 1992; 80(1):23-8.
2. Haynes RB et al. Computer searching of the medical literature: an evaluation of MEDLINE searching systems. *Ann Intern Med* 1985; 103: 812-6.
3. Peters TA. When smart people fail: an analysis of the transaction log of an online public access catalog. *J Acad Librarianship* 1989; 15:268-73.
4. Cahan MA GRATEFUL MED: a tool for studying searching behavior. *Med Ref Q* 1989; 8:61-79.
5. Fenichel CH Online searching: measures that discriminate among users with different types of experiences. *JASIS* 1981; 32:23-32.
6. King NS Search characteristics and the effects of experience on end users of PaperChase. *Coll Res Libr* 1991; 52:360-74.
7. Horowitz GL, Jackson JD, Bleich HL. PaperChase: self-service bibliographic retrieval. *JAMA* 1983; 250:2494-9.
8. Haynes RB et al. Online access to MEDLINE in clinical settings: a study of use and usefulness. *Ann Intern Med* 1990; 112:78-84.
9. Montgomery JD. Training of end user searchers: use and perceived benefits of online and non-online information resources. *Can J Inform Sci* 1987; 12:109.
10. Bleeker A et al. MEDLINE on CD-ROM: an evaluation after six month's use. *Online Rev* 1988; 12:200.
11. Hubble AB. MEDLINE access through an online catalog: a study of user reactions. *Proc ASIS Ann Meeting* 1988; 25:137-42.