Warner HR, Woolley FR, Kane RL. Computer assisted instruction for teaching clinical decision-making. *Comput Biomed Res* 1974;7:564-574.

Computer Assisted Instruction for Teaching Clinical Decision-Making

Homer R. Warner

Department of Biophysics and Bioengineering, University of Utah, Salt Lake City, Utah 84132

and

F. Ross Woolley and Robert L. Kane

Department of Family and Community Medicine, University of Utah, Salt Lake City, Utah 84132

Received February 10, 1974

A computer-assisted instructional program was devised to teach problem-solving behavior associated with history taking. The program focuses on the problem of identifying the information value of patient history questions. A total of 68 freshmen and sophomore medical students used the computer program in conjunction with their course on physical diagnosis. Each simulated case requires the student to make a series of tentative diagnoses which are compared against the most likely diagnosis as determined by the computer. The system operates in two modes, one with feedback for instruction, and a non-feedback method for testing.

Introduction

Practicing physicians are constantly called upon to apply clinical decision-making skills. The effects of such decision-making processes can have wide-ranging implications for the patient. In the extreme case, one can imagine that these decision-making skills will result in hastening or prolonging the death of the patient. Although most of the decisions made by physicians are not of such magnitude, decision-making necessarily constitutes one of the major activities in patient care. An examination of the skills to be acquired during medical education (1) suggests that problem solving or decision-making based on patient data is of paramount importance. As most medical school curricula are currently organized, decision-making is a process acquired secondarily in such experiences as clinical clerkships; there the student under the tutelage of an experienced physician must begin to assimilate his past experiences and information and apply them to the newly encountered patient situation.

Potentially more effective and efficient ways of organizing one's thinking in solving patient problems include such techniques as the application of formalized logic to specific clinical problems (2, 3). Ellstein (4) has shown that artificial encounters can be of value in teaching clinical-decision processes. In recent years, development of computer-assisted instructional techniques has proven to be of value in various parts of the medical school curriculum. A number of medical centers have become deeply committed to the use of computers in instructing medical students in various phases of their education. The University of Illinois (5) and

Ohio State University with their pilot medical school (6) are merely two of the more notable examples in which computers have been well integrated into the medical school curriculum. Coincident with these developments, the computer bas become a clinical tool for dealing with masses of patient data acquired through both inquiry and direct physical measurement.

The Model

A model of clinical decision-making for history taking and diagnosis was developed based on Baysean conditional probability (2). The set of 298 questions and 142 diseases covering all medical specialties has been defined and incorporated into a table which contains the probability that a patient with any one of the diseases will present with each of the symptoms. The initial estimates of these probabilities were made by the authors and a group of consultants in each of the special areas. Now that more than 12 000 patients have responded to these questions on admission to the hospital and have been ultimately diagnosed at the time of their discharge by the attending physician, actual frequencies for the occurrence of these symptoms in many of the more common diseases have replaced the initial estimates. With more time and experience with the less common illnesses, these numbers will be further improved.

As a student begins a teaching session at the terminal and requests a new simulated case to work on, the program randomly selects one of the diseases in the set using a random number generator. Having chosen a disease the program then scans the table for this disease across all symptoms. For each symptom the program generates a random number between 0 and 100 which is compared to the probability (in percent) that the symptom would occur in a patient with the disease chosen. If the random number selected is less than this probability, the simulated patient is considered to have that symptom. Thus, the pattern of symptoms in the simulated case may be different each time that disease is simulated, but the choice of symptoms is appropriately based on the probability of that symptom occurring in the disease. The “yes” answers so generated for a given simulated case are stored for future reference.

The program then presents back to the student the sex and age of the patient as just described and the primary complaint. The primary complaint is that symptom for which a “yes” answer was given which has the highest probability among all symptoms of occurring in a patient with this disease and for which this patient answered “yes” (see Fig. 1). This information is presented to the student along with a list of nine diseases and a tenth category labeled “none of these.” The diseases in this list are chosen using the following algorithm: a random number between 0 and 100 is generated, the list of probabilities by disease calculated from the Bayesean equation using the information from age, sex, and the chief complaint, is scanned. The probabilities are summed and as each new probability is added to the sum, it is compared to a random number. When the sum equals or exceeds the random number that disease is added to the list, thus even a disease whose

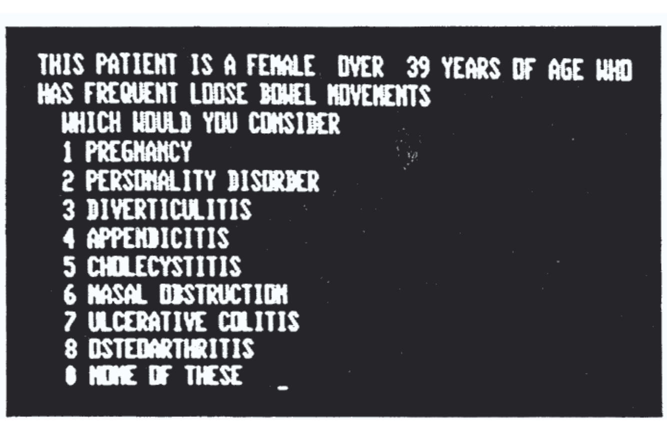


Fig. 1. Computer display of sex, age, and chief complaint of simulated case with possible diagnoses from which the student may choose one.

probability is 1 % may be in the list if it just happens to be the one which makes the sum equal or exceed the randomly chosen threshold. The student then chooses one of these diseases or selects the item "none of these."

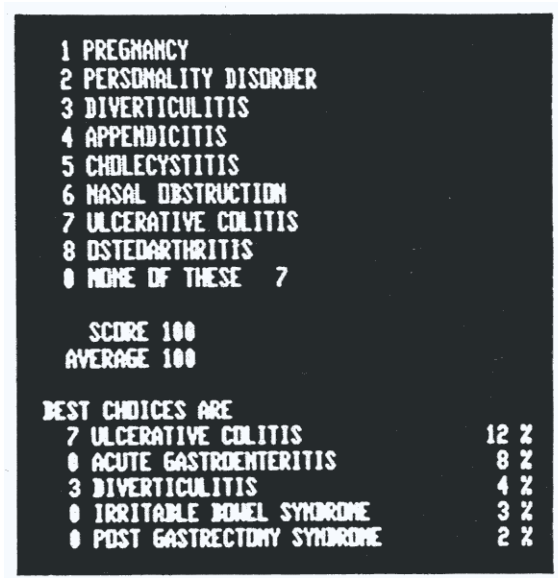


Fig. 2. The student chose item 7. The student's score is displayed along with the 5 most likely diagnoses calculated by the computer program.

At this point the student's choice is scored from 0 to 100 by taking one hundred times the ratio of the probability for the disease he chose to the probability for the best choice (the maximum probability in a set). If he chooses "none of these," the score is determined using the ratio of the highest probability among the diseases not in the list presented to the probability of the best choice. The program then presents to the student his score on this choice, the average of all his scores so far on this case, and a list of the five most likely diseases with the probability of each (Fig. 2). He then is presented with a list of 10 questions which have not already been asked and requested to select one, two, or three of these that would be most appropriate at this stage of the problem-solving process. The algorithm for measuring the appropriateness of a given question is based on the product of two terms A and B. The value of term A for a given question is obtained by calculating the probability for each

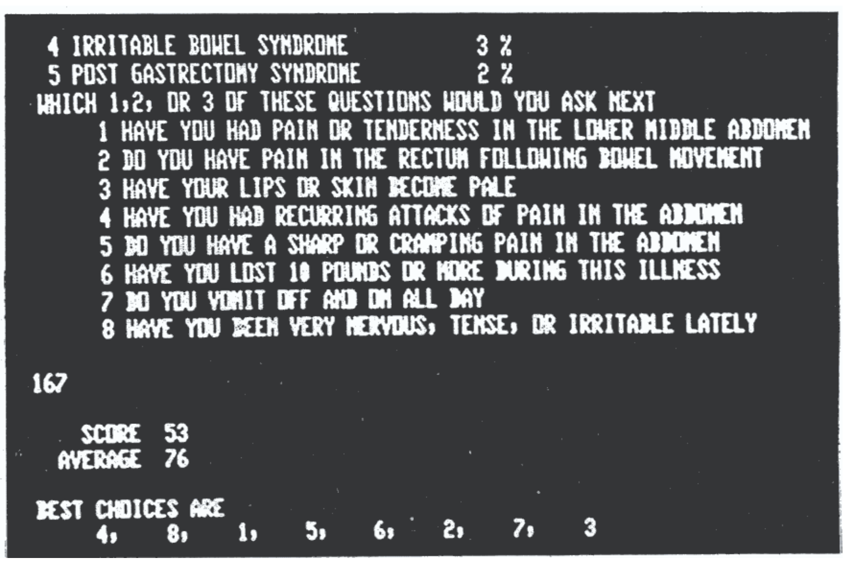


Fig. 3. From the set of questions the student chose items 1, 6, and 7 and received a score of 53. The choices are displayed in order of decreasing information content and probability of being answered yes at this stage of the problem.

disease were that question answered "yes" and subtracting the probability of that disease before asking that question. This difference is multiplied by the second term which is the probability of getting a "yes" answer to

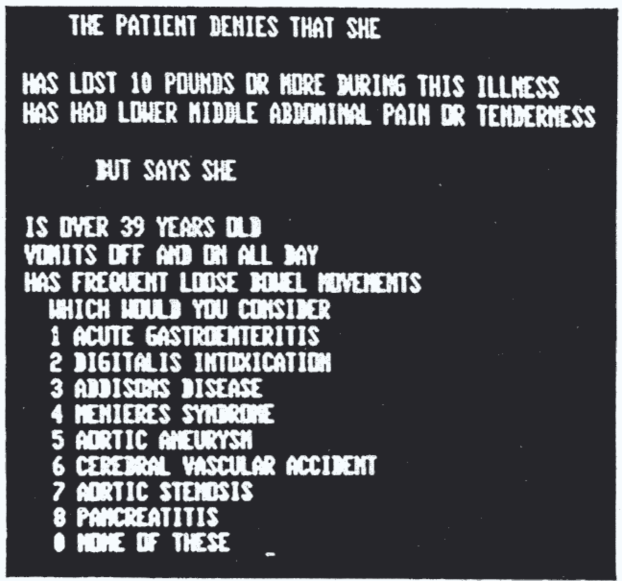


Fig. 4. The answers to the questions posed by the student are displayed and he is asked to choose again the most likely diagnosis based on this new information.

this question. This is obtained by summing across all diseases the probability of a "yes" answer to this question in each disease times the current probability of each disease.

Thus, a number is generated for each question which is a measure of its appropriateness and which may be incorporated into the summing algorithm described above for disease selection and used for selecting questions for presentation some of which will be appropriate and some inappropriate (Fig. 3). When a student selects one, two, or three of these questions to ask the theoretical patient, he is scored again by comparing the sum of the

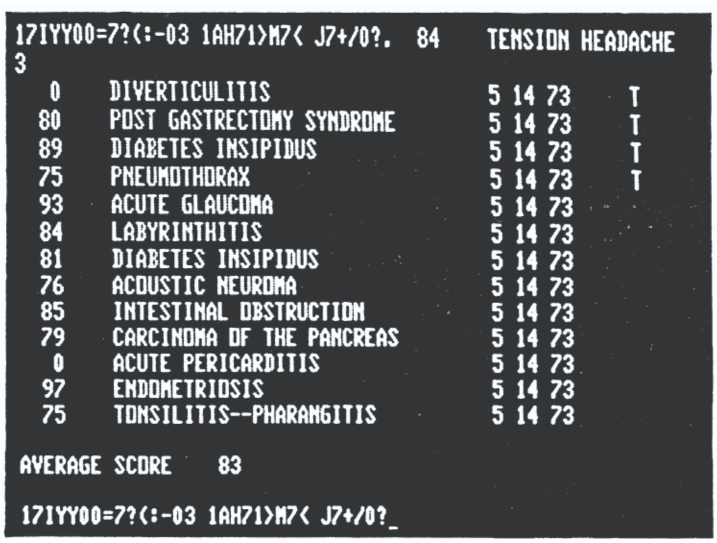


Fig. 5. A display of the diseases which have been simulated by the computer for this student the dates on which he saw each "case," and, in the left column, his score on each "case." A score of zero indicates this case was not completed and does not contribute to the average score.

measure of appropriateness for the question selected divided by the sum of the measures of appropriateness for the one, two, or three "best" questions.

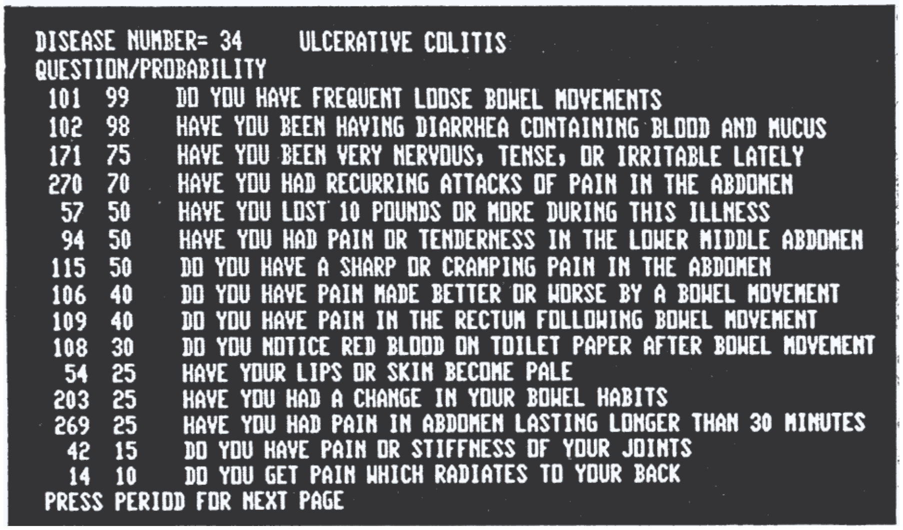
These questions are then "asked" the simulated patient and the answers previously stored are used appropriately in modifying the disease probability and are presented back to the student as shown in Fig. 4. The student again is presented with a list of diseases generated as described above from these new probabilities and is asked to select the one most likely at this stage in the problem-solving process. This cycle of events continues until either the disease probability for one diagnosis exceeds 90% or the appropriateness measure for the remaining unasked questions falls below a certain threshold.

Fig. 6. A display of the diseases most commonly associated with symptom 101. Item 1 indicates that 99% of patients with ulcerative colitis (disease 34) will answer yes to this question.

In addition to these features the program also has several other capabilities. One, it can be run in a test mode which operates exactly the same as in the teaching mode except that the correct answers, probabilities, and scores are not fed back to the student at each stage of the process. Two, a record is kept for each student of the cases he has seen, including the name of the disease, the score obtained, and whether or not the case was run in the test mode (Fig. 5). And three, an option which allows the student to review the probabilities of a particular symptom across all diseases listed in order of decreasing probability (Fig. 6) or for any disease across all symptoms listed again with the most probable symptom first (Fig 7).

Study Design

This study was designed basically as an evaluation of the computer-assisted decision-making process on early medical education, i.e., with freshman and sophomore students. Both internal and external factors were investigated as they affect the validity of such a teaching program. The internal questions were whether or not the students improved over time in their performance on the teaching program, the acceptance of the program by medical students, the ability to integrate such a program into already crowded curricula, and feedback from the participants as to the performance of the program. External questions which the study investigates relate to the students' performance in related course work as well as their performance in clerkship or clinical types of experience.



```
DISEASE NUMBER= 34     ULCERATIVE COLITIS
QUESTION/PROBABILITY
 101  99    DO YOU HAVE FREQUENT LOOSE BOWEL MOVEMENTS
 102  98    HAVE YOU BEEN HAVING DIARRHEA CONTAINING BLOOD AND MUCUS
 171  75    HAVE YOU BEEN VERY NERVOUS, TENSE, OR IRRITABLE LATELY
 270  70    HAVE YOU HAD RECURRING ATTACKS OF PAIN IN THE ABDOMEN
  57  50    HAVE YOU LOST 10 POUNDS OR MORE DURING THIS ILLNESS
  94  50    HAVE YOU HAD PAIN OR TENDERNESS IN THE LOWER MIDDLE ABDOMEN
 115  50    DO YOU HAVE A SHARP OR CRAMPING PAIN IN THE ABDOMEN
 106  40    DO YOU HAVE PAIN MADE BETTER OR WORSE BY A BOWEL MOVEMENT
 109  40    DO YOU HAVE PAIN IN THE RECTUM FOLLOWING BOWEL MOVEMENT
 108  30    DO YOU NOTICE RED BLOOD ON TOILET PAPER AFTER BOWEL MOVEMENT
  54  25    HAVE YOUR LIPS OR SKIN BECOME PALE
 203  25    HAVE YOU HAD A CHANGE IN YOUR BOWEL HABITS
 269  25    HAVE YOU HAD PAIN IN ABDOMEN LASTING LONGER THAN 30 MINUTES
  42  15    DO YOU HAVE PAIN OR STIFFNESS OF YOUR JOINTS
  14  10    DO YOU GET PAIN WHICH RADIATES TO YOUR BACK
PRESS PERIOD FOR NEXT PAGE
```

Fig. 7. A display of the symptoms most commonly associated with disease 34 (ulcerative colitis).

Two separate sets of subjects were used in this research; an experimental and control group were contained within each set. The first set of subjects consisted of freshmen currently enrolled in physical diagnosis in the University of Utah Medical School. The second set were sophomore students similarly enrolled in another course in physical diagnosis. Although the computer-learning program was not a required part of the class, there was a prior arrangement with each of the instructors that students were free to participate if they so desired Following a brief presentation of the mechanics of the computer-learning program followed by a short question-and-answer period, the students were invited to volunteer to participate in the experiment. The majority of students in both classes elected to participate in the experiment. During the presentation, however, students were informed that only half of those who signed up would be included as part of the experimental group. Once the lists were obtained, they were divided equally into experimental and control groups by a random number process. The latter were to be used for comparison of external validity measures. Thus the final numbers were made up of 37 experimental and 37 control for the freshmen, and 31 experimental and 31 control for the sophomores. Although a volunteer approach to participation has some biasing effect in terms of generalization to other classes, the effect of bias between experimental and control should be controlled by the randomization process used.

Procedure

The experimental groups for the freshmen and sophomores were called together independently. At this time they were each instructed as to how the computer actually worked and how they would be able "to play the game." The experiment was conducted over a period of six weeks, during which time each participant was required to play a minimum of ten games per week plus take two tests. As previously discussed, the only difference between games and tests was that during the game, the student received feedback as to the correctness of his approach, whereas during the test, no such feedback was given. As an additional stimulus, the students were paid one dollar for each test in which they scored over 85%. However, only two tests of over 85% were rewarded per week. A student could play as many games or take as many tests as he desired: however, his maximum pay was $2.00 per week. A prize of $25.00 was offered to the person in each class who achieved the highest over-all average test scores during the experiment.

Two terminals were used to insure availability These were connected over telephone lines with the computer located three miles away. The terminals were physically located in the basement of the Medical Library along with other computer terminals which were connected to computer learning programs at other universities. Access to these terminals was restricted by student code number known only to the student and by a check-out device which was necessary to power the computer terminal. This check-out device was controlled by library personnel and given only to authorized users. No students other than the experimental groups from freshman and sophomore classes were permitted to use this computer terminal.

To obtain feedback on performance, a log was maintained at the same place where the activation device was checked out. During the six weeks of the experiment, some 26 entries were made onto this log, reflecting various problems which were encountered in the operation. Scores achieved by the students on the test were stored automatically in the computer and printed out on a weekly basis. In addition, the number of games played by each student as well as the assessing of games by disease type were also maintained in order to provide a check as to how many of the diseases were being presented and a relative score as to how well students were doing on each of these diseases. An emergency phone number was also provided so that students could obtain rapid information when problems were encountered with the terminal devices.

Results

During the six weeks of the experiment, 7058 cases were reviewed by the 68 experiment students or an average 104 cases per student. The average score for 68 participants ranged from 91 for a student who saw over 400 cases, to 74 for a student who saw 59 cases. Table 1 shows the incremental change from the first week to the last week in terms of the improvement of the average score. On the first 20 cases, as compared to the last 20 cases, over-all, sophomores increased by three points, from 82 to 85, and the freshmen increased by three points from 80 to 83. Although not statistically significant, the higher average initial score for sophomores is consistent with their relatively greater understanding of clinical medicine than for the freshmen. The internal

TABLE 1
Comparison of Control and Experimental Group Scores[a]

	N	Scores, first week	Scores, last week	Final exam score		Preceptor rating	
		Average	Average	Average	S.D.	Average	S.D.
Freshmen							
experimental	37	80.32	82.68	56.7	5.66	2.216	0.672
control	37	-	-	54.5	6.92	2.189	0.672
Sophomores							
experimental	31	82.00	84.33	23.86	5.41	2.258	0.893
control	31	-	-	24.87	5.89	1.871	0.670

[a] All differences between experiment and control groups not significant at $P < 0.05$,

validity of the program seems to have been demonstrated by the students' consistency in taking tests and continued participation in the experiment. Anecdotally, it must be noted that many students approached the authors during the summer asking if the computers might be available again during the coming school year.

The matter of external validity cannot be completely answered at this time. A series of statistical analyses were performed to try to determine what effects, if any, the computer program may have had on the experiment students' performance in their class as compared to that of the control groups. The dependent variables used in this measure were final examination scores for each of the two classes, as well as the somewhat more subjective scores of those who acted as preceptors for the students during their physical diagnosis class. No significant differences could be determined by any of the analyses performed. It should be noted, however, that although no improvements might have come directly to performance on the examinations, a careful analysis of these exams indicated that the test material was somewhat different in content than the nature of the computer learning program. Consequently, the lack of statistical significance in this case is probably inconsequential. Continual analysis will be conducted to try to determine what effects the computer learning program has on decision-making of students as they move into their clinical experiences. Subsequent evaluations will be made on their performance through a variety of clinical rotations. In addition, it must be noted that the students voiced considerable enthusiasm for continuation of this kind of independent instruction.

The role of such a program as a device for testing problem-solving skills in an objective fashion also holds real promise. Expansion of the program to include data from laboratory and physical examinations is underway and should add significantly to its value as a tool for teaching clinical decision-making skills.

Summary

The use of a computer-based learning device to assist in the learning of clinical decision-making cannot be fully evaluated from the foregoing experiment. Additional data must be collected to validate the long-term results of such an experience. There is evidence, however, that improvement can be made in the ability of students to "agree" with the computer and its evaluation of various physical symptoms. A variety of factors, including improvement of knowledge of medical terminology, and a satisfaction in having some quasiclinical experience identifying patient problems, all have merit in supporting continued activity with this type of teaching program.

References

1. Kane, R. L., Woolley, F. R., and Kane, R. Toward defining the end product of medical education. *J. Med. Educ.* **48,** 615-62 (1973).
2. Gorry, G. Anthony and Barnett, G. Octo. Experience with a model of sequential diagnosis. *Comput. Biomed. Res.* **1,** 490-507 (1968).
3. Warner, Homer R., Rutherford, Barry D., and Houtchens, Bruce. A sequential Bayesean approach to history taking and diagnosis *Comput. Biomed. Res.* **5,** 256-262 (1972).
4. Ellstein, A. S., Kagan, N., Shulman, L. S., Jason, H., and Loupe, M. J. Methods and theory in the study of medical inquiry. *J. Med. Educ.* **47,** 85-92 (1972).
5. Schorow, M. Problem solving theory and the practice of clinical medicine. *Con. Med. Ass. J.* **97,** 711-716 (1967).
6. Prior, J. A., Griesen M. A., and Folk, R. L. Computer-assisted independent study: A pilot program. *J. Med. Educ.* **45,** 801 (1970).