

# A Randomized Comparison of Alternative Formats for Clinical Simulations

Charles P. Friedman, Ph.D., Cynthia L. France, Ph.D., Douglas A. Drossman, M.D.

Laboratory for Computing and Cognition  
The School of Medicine  
University of North Carolina at Chapel Hill  
Chapel Hill, NC 27599-7530

## Abstract

Computer-based clinical simulations for medical education exhibit wide variation in structure and format, yet few studies have examined which formats are optimal for particular educational settings. This study is a randomized comparison of the same simulated case offered in three formats: a "pedagogic" format offering explicit educational support, a "high fidelity" format attempting to model clinical reasoning in the real world, and a "problem solving" format that required students to express specific diagnostic hypotheses. Data were collected from rising third year medical students using a posttest, attitudinal questionnaire, students' writeups of the case, and log files of students' progress through the simulation. Student performance on all measures differed significantly by format. In general, students using the pedagogic format were more proficient but less efficient. They acquired more information but were able to do proportionately less with it. Students at this level had difficulty generating queries for clinical information using natural language. The results suggest that the format of computer-based simulations is important educationally; a format that works well for one learner population may not work as well for another population with different characteristics.

## Introduction

Interest in educational use of computer-based clinical simulations took root in the early 1970s. The most visible products of this early work included the CASE system developed at the University of Illinois (1) and a series of simulations developed at the Massachusetts General Hospital (2). At about the same time, Richard Friedman built simulations into the fabric of clinical teaching in an internal medicine clerkship (3,4). Friedman (4) noted several potential advantages of using simulations in clinical education:

- 1) All students are given experience with a core group of disorders regardless of patient population or service in the hospital...
- 2) Students are given an opportunity to see the importance of cost and test availability in determining what tests are truly most valuable.
- 3) Since all students can work up the same case, simulation is the ideal vehicle for teaching diagnostic skills.
- 4) Students are given detailed self-evaluation feedback...so that emerging deficiencies can be rapidly corrected.
- 5) Mistakes are made on a computer and not a patient.

[p. 829]

\* Current Affiliation: Digital Equipment Corporation, Bedford, MA

Several design variants were evident in these simulation programs. Some were "static" simulations, providing for students the opportunity to work up a simulated case only at one point in time, while other, "dynamic" simulations allowed students to experience the time evolution of the case in response to their own management decisions. The CASE system required students to obtain information about the "patient" using natural language query instead of choosing menu items.

Computer-based simulations using mainframe technology did not, in general, disseminate beyond the institution where each program was developed (2). One possible reason was cost on a student-hour basis which, despite promising early projections, turned out to be much higher than comparable costs of classroom instruction (5).

With the advent of relatively powerful and inexpensive microcomputers in the early 1980s, the number of computer-based systems for patient simulations proliferated rapidly (6-9). These systems generally employed the dynamic model whereby students could specify treatment of the simulated "patient" and encounter the patient again at a later point in simulation time. However, the mixed modality of user-interfaces remained: some systems using natural language, others using menus. The origin of this diversity was rather clear. Some system developers would argue that the work required to produce natural language simulations was justified by the element of realism they introduce, since experienced physicians do not bring checklists of interview questions to their encounters with real patients (10). Others would argue that the cognitive task of selecting items from a very long menu is equivalent to a natural language environment, since it is very inefficient to "shop through" a lengthy list.

Systems also varied in the amount and timing of feedback they presented to learners. Some simulations provided directive feedback during the simulation, keeping students generally on track in the management of the patient (6). Other simulations followed more closely the "flight simulator" model by providing relatively little explicit feedback, letting the realistic clinical consequences of clinical decisions indicate to the student whether he/she was proceeding on a fortuitous path (8).

There also developed serious interest in using simulations for formal assessment of clinical skills. The CBX (Computer Based Exam) project was initially sponsored by the American Board of Internal Medicine and later adopted by the National Board of Medical Examiners (11,12). Current plans call for the introduction of CBX into Part III of the National Board examination early in the 1990s.

At this writing a new generation of simulation programs for medical education is under development. An obstacle to the proliferation of simulations in medical education has been a relative paucity of case material for students to employ. In most systems, each case must be authored essentially from scratch, a labor-intensive task that falls primarily to an expert

clinician. To address this problem, the ILIAD system (13), when functioning in a simulation mode, can generate medically plausible cases probabilistically from its knowledge base. Another perceived drawback of educational simulations has been the lack of an explicit pedagogic strategy in the programs. That is, most simulations do not direct students to employ a systematic approach to clinical problem-solving. A system developed at Southern Illinois University invokes an explicit hypothetico-deductive framework which compels students to think systematically through a clinical problem by asserting a diagnostic hypothesis before eliciting data to test that hypothesis (14).

### Structure of the Study

As the number of clinical simulation systems continues to increase, a great deal of diversity in their format remains. This study explicitly compares alternative formats for these programs, exploring which is preferable along several dimensions of educational outcome. The approach we selected was to give random samples of students simulations of the same clinical case, but with systematically varied format. The three formats used were based on approaches that are prevalent among existing simulation systems:

- A "pedagogic" model that provided explicit educational support. This version gave students access to clinical information via hierarchical menus and provided educational feedback as the simulation unfolded.
- A "high fidelity" model that attempted as closely as possible to mimic clinical reasoning in the real world. Students requested information in natural language and received no feedback other than the clinical consequences and monetary costs of their decisions.
- A "problem solving" model that imposed a hypothetico-deductive approach by requiring students to assert a working hypothesis before eliciting data, and required them to assert whether each item of information received tended to confirm, rule out, or uninformative regarding the stated hypothesis.

We compared student performance on a number of educational outcomes, with emphasis on information access and recall. This emphasis was directed by the importance of problem-specific knowledge in clinical reasoning (15,16). In exploring these outcomes, we are interested in issues of proficiency (How much critical information do students access and subsequently recall?) as well as efficiency (What proportion of information that students access and recall is critical information?) (17). Multiple outcomes assessed by multiple measurement tools are employed to create a more comprehensive picture of each simulation format's effects than would derive from just one outcome measure.

### Investigative Questions

Specifically, this study explores whether students employing alternative formats of clinical simulations differ in:

- 1) Behavior on the Simulation Itself:
  - a) the amount of clinical information accessed while working through the case;
  - b) the amount of critical clinical information accessed while working through the case;
  - c) the proportion of clinical information accessed that is critical information.

### 2) Content of a Case Write-Up:

- a) the total amount of information about the case noted in the write-up;
- b) recall of information about the case: the proportion of information accessed during the simulation that is entered in a case write-up;
- c) the amount of information inferred: information noted in the case write-up directly accessed during the simulation;
- d) the ratio of critical information reported in the write-up to that accessed while working the case.

### 3) General knowledge about the case as reflected in scores on a posttest.

### 4) Perceived value of the simulation as an educational experience.

### Methods

*Design:* The study employed a randomized, three group design. The design had one independent variable, format of the simulation, with three levels of this variable corresponding to the three discrete simulation formats employed. Subjects were randomly assigned to one of the three simulation formats. Each of nine outcome measures constitutes a separate dependent variable. The outcome variables were assessed by analysis of log files maintained by the computer system, and by asking each subject to complete a "write up", a posttest, and a questionnaire. To add precision to the study, we also explored the effects on the outcome measures of prior achievement in basic science courses relevant to the simulation content.

*Subjects:* Subjects in the study were rising third year medical students at the University of North Carolina (UNC) School of Medicine. The simulation exercise was a formal part of the "Transition" course within the UNC medical curriculum, designed to bridge students from the pre-clinical to the clinical phases of their education. Of the 145 medical students enrolled in the course, eighty participated by administrative assignment in the computer simulation exercise while the remaining students studied a different topic. Participation in the exercise was an integral part of the course, so students were motivated to perform well. Complete data for use in the study were obtained from 72 students.

*Simulation Program:* Students worked through all three versions of the simulations using author/driver software developed for Apple II series computers by one of the authors (CPF). In its generic form, this software contained many features needed for this research, including routines for interpreting natural language queries and a log file feature that recorded student trajectories through the simulation. Specific variations were made in the driver program as needed to create the alternative presentation formats for the common case script. In earlier trials, students reported the program to be acceptably fast and accurate in interpreting natural language requests for information. Because the software had been used by two previous cohorts of students in the Transition course at UNC, it was essentially free of bugs and its mechanics of use had been improved each year.

The simulations, in all three formats, had a node and branch structure. Students encountered the "patient" in a series of scenes. During each scene the student could elicit history, physical exam, and lab/test data. When the student completed data elicitation in a given scene, he/she was presented with a list of up to seven discrete management

options. The option selected determined the scene to which the student was branched, corresponding to a realistic clinical change in the patient's condition. When the student branched to a new scene, all patient data were updated accordingly. When branching to a new scene, the student was told the cost of care for that scene and total costs incurred in the simulation to that point.

The three formats used in this research were variants on this general scenario. The manner in which data were elicited within each scene of the simulation depended on the format of the simulation. Natural language query--with students typing in their requests--was employed in the "high fidelity" and "problem solving" versions. Data elicitation in the "pedagogic" version was based on a hierarchy of menus; students made choices from given lists. (For example, available history items were divided into eight sub-menus from which students could make specific selections.) In the "pedagogic" version only, students could receive feedback on the plausibility of a number of diagnoses at the end of each scene. In the "problem-solving" model, students had to assert a working hypothesis before eliciting any data and also were required to assert the impact of each data item received on their evaluation of the working hypothesis. The working hypothesis could be changed at any time.

*The Clinical Case:* The simulated case employed in the study ("Mrs. Jones") involves a 57 year old woman who presents with abdominal pain, claiming that her previous physicians have been unable to help her. Mrs. Jones has six problems that could be diagnosed separately and a seventh problem (urinary tract infection) that develops only if the student elects to hospitalize her at a particular point in the simulation. Despite the complexity of the case, the two most important diagnoses of diabetes mellitus and depression are not difficult to reach. When used in the same course during the two previous years, this case problem was found to be an appropriate, though very challenging, educational exercise for rising third year students. The total case consisted of nine discrete scenes. A total of 127 history, exam, and lab items was available in each of the scenes. The author of the case had earlier designated 43 of these items as "critical" to a complete work-up.

*Research Instruments:* Data to address the research questions were collected using several methods. We were able to record, for later analysis, the data items accessed by each student during the simulation using the log file feature of the simulation driver program. To address students' unstimulated recall of case information, we used a hospital progress note form on which each student recorded a "write up" of the case. We assessed students' general knowledge about the case using a 22 item posttest developed collaboratively by the authors. The posttest covered the interpretation of critical data items about the patient, the appropriateness of the important diagnoses and the information supporting them, and management of Mrs. Jones' many problems. Students' satisfaction with the simulation was addressed using a 10 Likert scale items on a self-administered questionnaire. Relevant prior achievement was assessed using students' scores on the endocrinology and psychiatry course examinations in the second year medical curriculum at UNC. These courses were chosen because their content was basic to the primary diagnoses in the simulated case.

*Procedures:* As a part of his/her schedule for the Transition course, each participant in the study was assigned one afternoon (of three) to work the simulation, complete the study instruments, and attend a group discussion of the case led a medical preceptor. Students had been previously randomized into the treatment groups. When each student arrived for the session, he/she was given a diskette containing the simulation in the pre-assigned format. Students were told

they had approximately one hour to work individually through the case. The authors were available to the students to help with system mechanics and to answer questions of medical terminology. After finishing with the simulation, students brought their diskette (which contained their log files) to a member of the staff and received the write-up form, the posttest, and the questionnaire. When students completed these forms they reported to a classroom for the group discussion.

*Analysis:* Complete data were acquired from 72 students: 26 in the "pedagogic" group, 24 in the "high fidelity" group, 22 in the "problem-solving" group. Each student's log file was analyzed for data items accessed and the proportion of those items that were "critical." One of the authors read all case write-ups and coded these for the presence/absence of each data item. Since this process required some subjective decisions, a second coder crosschecked a representative sample of eight write-ups. Agreement coefficients among the coders (Cohen's Kappa) averaged .85 for these items, pointing to high reliability of these data. The posttest, in multiple-choice format, was scored based on total number of correct responses. Items on the questionnaire were analyzed individually and responses to ten items were summed to compute an index of satisfaction.

Early in the data analysis, we found that the measures of prior achievement had no significant relationship to the outcome variables in the study ( $r < .12$  for all variables). Data were subsequently analyzed using one-way analysis of variance to explore global differences among the simulation formats. Separate analyses were employed for each outcome variable. Scheffe's test was used to explore pairwise differences among formats in instances where a global difference was found using an F-test. The Kruskal-Wallis nonparametric test was used to explore group differences on individual questionnaire items.

## Results

Table 1 summarizes the findings for the major dependent variables in the study. The table reports means and standard deviations for each group, the results of the analysis of variance (F-test) to test global differences between the groups, and the results of Scheffe's test exploring differences between pairs of groups. From the Table 1, it is evident that the format of the simulation was related to all dimensions of learning explored in this study.

Students using the "pedagogic" version of the simulation, which makes information accessible from menus, did accumulate a much larger set of clinical information about "Mrs. Jones" than users of the other two versions ( $p < .05$  by Scheffe's test) who obtained data using natural language. The same finding was observed for numbers of critical information items that were accessed, which is an index of proficiency. However, a different result obtains for the efficiency index which is the proportion of items obtained that were critical items. Here we observe higher efficiency for the "high fidelity" and "problem-solving" formats ( $p < .05$  by Scheffe's test). Since 43 of 127 data items (34%) were designated as critical, the efficiency index of students in the "pedagogic" group (.38) suggests that they were essentially indiscriminating in their search for information. An efficiency index of .34 would be expected by chance alone.

With regard to the case write-up, students using the "pedagogic" version listed more items of information than did users of the other two versions ( $p < .05$  by Scheffe's test). As evidenced by the recall index, students using the "high fidelity" and "problem-solving" versions of the simulation were able to recall in their case write-ups a higher proportion of what they had accessed. The inference index addresses the number data items that were not explicitly accessed during the

**Table 1:**  
Means, Standard Deviations and Tests of Significance by Format

Outcome Measure	Mean (SD) for Each Format*			Test of Group Differences	
	Pedagogic	High-Fidelity	Problem-Solving	F(2/69)	Significance
Total Items Accessed	58.65 (19.49)	<b>25.25</b> (8.77)	<b>21.36</b> (6.87)	58.93	p<.001
Critical Items Accessed	22.00 (6.19)	<b>11.96</b> (3.03)	<b>10.45</b> (3.22)	47.08	p<.001
Efficiency Index	0.38 (.06)	<b>0.48</b> (.09)	<b>0.51</b> (.10)	15.52	p<.001
Total Items in Write-Up	18.57 (9.15)	15.08 (4.98)	11.35 (5.45)	6.22	p<.01
Recall Index	0.27 (.17)	<b>0.43</b> (.14)	<b>0.38</b> (.23)	6.15	p<.01
Inference Index	<b>2.77</b> (1.86)	4.29 (1.85)	<b>3.14</b> (1.94)	4.37	p<.05
Critical Recall Ratio	0.39 (.18)	<b>0.56</b> (.18)	<b>0.54</b> (.21)	6.51	p<.01
Posttest Raw Score	13.7 (2.52)	<b>11.6</b> (2.22)	<b>11.4</b> (2.91)	6.11	p<.01
Satisfaction Index**	30.52 (5.94)	<b>37.42</b> (6.22)	<b>34.41</b> (8.08)	6.70	p<.01

\*Means listed in bold face do NOT differ significantly by Scheffe's test at the .05 confidence level.

\*\*Lower scores imply higher levels of satisfaction.

simulation but were listed explicitly in the case write-up. Because the write-ups were completed immediately following work at the computer, it is necessary that students inferred these items from the information they did access while working the case. By Scheffe's test, users of the "high fidelity" version inferred more information than users of the other two versions. The critical recall ratio indicates the proportion of all critical items accessed in the simulation that also appeared in the write-up. Students using the "problem-solving" and "high fidelity" versions had significantly higher critical recall ratios than students using the "pedagogic" version.

The posttest addressed the major diagnostic and management issues in the case and, as such, is a broader indicator of overall performance and understanding of the case than the information access and recall variables described previously. The average score for all students on the posttest was 12.2 (55.5%). As seen in Table 1, students in the "pedagogic" group outscored the other students.

An overall satisfaction index was computed by summing 10 Likert-scale items in the questionnaire. As indicated by the lower mean score in Table 1, the "pedagogic" format was rated most highly by the students. Differences between the "high fidelity" and "problem solving" formats were not significant. Table 2 presents mean ratings for each of the 10 questionnaire items. The items that differed significantly across the formats are indicated in bold type. The "pedagogic" format received significantly superior ratings for four items relating to information access and understanding of the case,

by the Kruskal-Wallis non-parametric test. The "high fidelity" version received superior ratings on realism.

### Discussion

These results lead to an overarching conclusion that simulation format makes a difference educationally. Students responded to these alternative formats in systematically different ways, largely in accord with theoretical and intuitive expectations. It follows that designers of clinical simulations and faculty members planning to employ them in medical curricula should consciously tailor their choice of format to their specific objectives and the backgrounds of their students.

Students using the "pedagogic" version accessed more items of data and more critical items of data, but were able to do proportionately less with them. These students were more proficient but less efficient. Students using the "pedagogic" version scored higher in the posttest, but not in proportion to the amount of information they acquired about Mrs. Jones as her case unfolded. The menu format made it easier for them to access information, and the questionnaire revealed that they appreciated that feature, but they did seem to "window shop" for data--seemingly acquiring information because it was there rather than because it was necessary. Students in the "problem-solving" group obtained the least information overall, and the least amount of critical information. The overhead of asserting hypotheses and evaluating each data item against the current working hypothesis forced students to consider clinical data more thoroughly than the other students were compelled to do. Given more time than students in this study were provided,

**Table 2:**  
**Mean Ratings for Likert-Scale Questionnaire Items**

Item*	Format**		
	Pedagogic	High-Fidelity	Problem-Solving
Program was Easy to Use	1.85	2.42	1.86
Program Held My Interest	1.77	2.00	1.59
<b>Taught Organization of Clinical Data</b>	1.89	2.58	2.18
<b>Helped Relate Data to Diagnoses</b>	1.92	2.58	2.22
<b>Helped Me Think Systematically</b>	1.88	2.46	2.27
Matches the Way I Think	2.08	2.71	2.41
<b>Was a Realistic Experience</b>	2.74	2.02	2.43
Difficult for Student at this Level	2.73	2.37	1.86
I Learned About Clinical Reasoning	1.92	2.08	2.00
<b>Format Made Data Access Difficult</b>	2.81	1.88	2.14

\*Items listed in bold face differed significantly by format ( $p < .05$  using Kruskal-Wallis test).

\*\*Responses on scale from "1" (strongly agree) to "4" (strongly disagree).

and students' willingness to spend more time on a simulated case, the "problem-solving" model may yield educational benefits untapped by the design of this study.

Students using the "high fidelity" and "problem solving" versions of the simulation reported difficulties in accessing the information they felt they needed to work through the case. There are several possible origins of this sentiment. Students at this level may not know what to ask for, or how specifically to ask for it. These difficulties would have been exacerbated by the occasional inability of the program to correctly interpret students requests as initially entered. When this occurred, the students may not have a readily accessible set of synonyms to employ to ask the question in a different way. A "help" feature that suggests synonyms would have been useful to these students.

These results must be interpreted in light of the educational level of the subjects and generalized with great caution to trainees at other levels. These rising third year students were just beginning their learning of clinical medicine. The features of the "pedagogic" simulation format--particularly the hierarchical menus--eased their data acquisition and helped them learn the structure of clinical information. This structure is something a fourth year student might take for granted, but it is a novel concept to students at early stages of clinical training. The findings of this study--particularly those findings relating to efficiency--suggest that students may derive educational benefit from constructing requests for information rather than choosing from a list. Overall, however, students

who are clinical neophytes are probably best served by a simulation that offers a great deal of support in obtaining clinical information and explicit educational feedback on various diagnostic and management options. It is not clear how quickly students in active clinical training would outgrow this format as they progress in clinical training. Both alternatives we studied--the "high fidelity" and "problem-solving" formats--have features that seem attractive for trainees at higher levels and for continuing education of physicians in practice.

### References

- [1] Harless, WG, Drennon, GG, Marxer, JJ, Root, JA, Miller, GE. CASE: A Computer-Aided Simulation of the Clinical Encounter. *J. Med. Educ.*, 46:443-448, 1971.
- [2] Hoffer, EP, Barnett, GO, Farquhar, BB, and Prather, PA. Computer-Aided Instruction in Medicine. *Ann. Rev. Biophys. Bioeng.* 4:103-118, 1975.
- [3] Friedman RB. A computer program simulating the patient-physician encounter. *J. Med. Educ.*, 48: 92-97, 1973.
- [4] Friedman, RB, Korst, DR, Schultz, JV, Beatty, E, Entire, S. Experience with the Simulation Patient- Physician Encounter. *J. Med. Educ.*, 53:825-830, 1978.
- [5] Kopstein, F.F., and Seidel, R.J. Computer-assisted Instruction Versus Traditionally Administered Instruction: Economics. *Audio Visual Communication Review*, 2:147-175, 1968.
- [6] Abdulla, AM, Watkins, LO, Henke JS. The use of natural language entry and laser videodisk technology in CAI. *J. Med. Educ.*, 59:739-45, 1984.
- [7] Whiteside MS, Lang NP, Kemp FF, Berry DH. Almost painless: using MED-CAPS to develop a CAI program. *Focus on Surgical Education*, 6: 8-10, 1988.
- [8] Friedman CP, Beery MP. Clinical simulations emphasizing psychosocial aspects of patient care. *Innovations in Medical Education*. Washington DC: Association of American Medical Colleges, 1983. (Abstract)
- [9] Schwartz MW, Hanson CW. Microcomputers and computer-based instruction. *J. Med. Educ.*, 57:303-7, 1982.
- [10] Feightner, JW, Norman, GR. Computer based problems as a measure of the problem-solving process: Some concerns about validity. *Annual Conference on Research in Medical Education*, 17: 51-56, 1978.
- [11] Update on computer-based testing. *National Board Examiner* 34 (1): 1-4, Winter 1987.
- [12] Orr NA. Non-cognitive correlates of performance on conventional and interactive components of a computerized examination. *Annual Conference on Research in Medical Education*, 27: 284-289, 1988.
- [13] Warner HR, Haug P, Bouhaddou O, et al. ILIAD as an expert consultant to teach differential diagnosis. *Proceedings of the 1988 Symposium on Computer Applications in Medical Care*, pp 371-376, 1988.
- [14] Williams, RG, Vu, NV, Barrows, HS, and Verhulst, SJ. The clinical reasoning test: An objectively scorable measure of clinical reasoning. In I.R. Hart, R.M. Harden & H.J Walton (Eds.). *Newer Developments in Assessing Clinical Competence*, Ottawa, Canada: Heal Publications Ltd., 1986
- [15] Feltovich, PJ, Barrows, HS. Issues of generality in medical problem solving. In M.C. Devolder and H.G. Schmidt (Eds.). *Tutorials in Problem Based Learning*, Assen, Holland: Van Gorcum, 1984.
- [16] Elstein, A, Shulman, LS, Sprafka, SA. *Medical Problem Solving*. Cambridge: Harvard University Press, 1978.
- [17] McGuire CH, Solomon LM, Bashook PG. *Construction and Use of Written Simulations*. New York: Psychological Corporation, 1976.