Computer-based Testing in Family Practice Certification and Recertification

Michael D. Hagen, MD, Walton Sumner, MD, Guy Roussel, MD, Richard Rovinelli, PhD, and Jinzhong Xu, PhD

Background: **The member boards of the American Board of Medical Specialties have agreed to expand the scope of certification to include assessment of medical knowledge, practice-based learning and improvement, patient care, interpersonal and communication skills, systems-based practice, and professionalism. Multiple-choice examinations provide limited ability to assess these dimensions.**

   Methods: **The American Board of Family Practice (ABFP) has developed a computer simulation system to facilitate more comprehensive candidate evaluation. The system consists of a knowledge base, a simulation program to create patient scenarios, an interface for presenting simulations to users, and an administrative database to track candidate performance and interactions with the system. The system uses population distributions for disease states to produce cases and evolves patients in response to candidate interventions, such as pharmacological and nonpharmacological therapies. We use Bayesian belief networks to model patient characteristics and comorbid condition interactions.**

   Results: **Simulations have been created for 7 disease states; ultimately simulations will be available for 25 to 30 disease states. Initial testing will take place in regional examination centers but will ultimately use the Internet for convenient access for certification and recertification candidates.**

   Conclusion: **The ABFP will begin field-testing the system in early 2003 and will include simulations in the certification and recertification examination process in 2004. (J Am Board Fam Pract 2003;16: 227–32.)**

The medical specialty boards have recently expanded their perspective on physician performance to include six components: medical knowledge, practice-based learning and improvement, patient care, interpersonal and communication skills, systems-based practice, and professionalism[1] (detailed descriptions of these components can be found at http://www.acgme.org/outcome/comp/compFull.asp.) Traditional formats such as multiple choice examinations focus on knowledge assessment. The additional dimensions, however, will require alternative techniques, such as standardized patients, chart audit, objective standardized clinical examinations, and simulated patient cases.[2]

   The large number of candidates the American Board of Family Practice (ABFP) examines annually renders standardized patients and objective standardized clinical examinations logistically impractical. Limited chart audits are currently included in the comprehensive office record review, conducted as part of the recertification process. Various groups have experimented with clinical simulation since the 1970s. Barnett and colleagues[3,4] developed case simulations at the Massachusetts General Hospital, and developers at the University of Illinois created the CASE (Computer Associated Simulation of the Clinical Encounter) at approximately the same time.[5] The American Board of Internal Medicine experimented with this system for possible use in physician certification and found this approach to be feasible.[6] Another system designed at the University of Wisconsin served as the basis for the National Board of Medical Examiners Step III licensure examination,[7,8] and their experience provided further validation for the role of simulation in professional assessment.[9]

   These systems use authored case descriptions that, once used, are no longer secure, necessitating a continual process of new case generation. The ABFP recognized in the early 1990s that simulation could improve their ability to make certification decisions but desired to avoid shortcomings associated with previous efforts. The Board established a
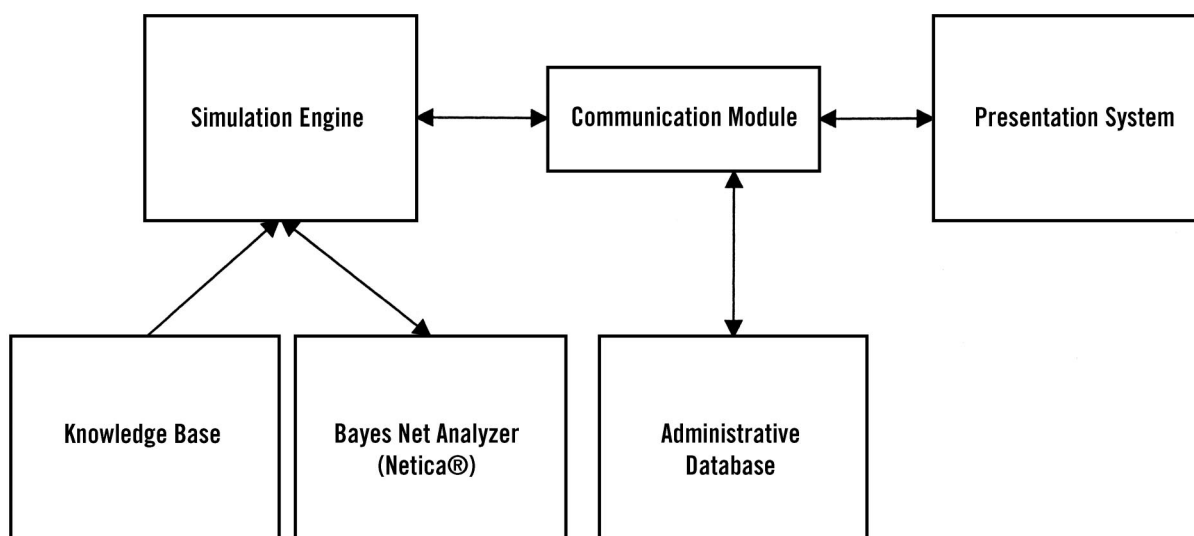
Figure 1. Schematic of system architecture for case simulation. The four main components consist of the knowledge base, the simulation engine, the presentation system, and the administrative database. The system also uses a third party package (Netica, Norsys Systems) for belief network analysis and manipulation (a fully functioning demonstration version of the software is available at www.norsys.com.) The simulator communicates with the presentation system through the communication module. The presentation system can interface with the outside world by means of the Internet or within a local intranet structure, such as at a test center.

development team that included computer scientists, mathematicians, psychometricians, and family physician informaticists. After examining existing architectures, the team adopted dynamic simulation,[10] driven from a knowledge base, as the means to create unique patient care scenarios appropriate to the family practice domain. This architecture creates unlimited similar, but unique, case scenarios for use in the certification process. The development team described the model architecture in earlier reports,[11,12] and the ABFP has patented the technology (US patent #6,246,975).

Several assumptions motivated the development effort:

1. The simulator should generate unique scenarios for each candidate, thus minimizing security considerations associated with the current paper examination.
2. The simulator should create scenarios from a knowledge base defined by family physicians and appropriate domain experts.
3. The scenarios should be realistic and evolve patients' health states prospectively with time and in response to candidate interventions.
4. The presentation format should provide a friendly interface that depends as little as possible on candidate computer skills.

5. The system architecture must provide structures suitable for a valid and reproducible scoring process.

Methods

The developers have described the architecture in detail elsewhere.[11–16] The simulator currently includes four fundamental pieces: a knowledge base, a simulation engine or program, the presentation interface, and an administrative database (Figure 1). The system uses population distributions for disease states to produce cases and evolves patients in response to candidate interventions. Bayesian belief networks model patient characteristics and comorbid condition interactions.

Knowledge Base Development

The knowledge base consists of a software program designed to support entry and maintenance of information (an object-oriented database[17] developed for the project) and the actual information needed to support simulation. The development team has created the knowledge base using knowledge development teams. Each knowledge development team consists of several family physicians (at least one of whom must be a current or recent member

of the ABFP Board of Directors), a domain expert (such as a specialist or family physician with special expertise in the particular domain), and a knowledge engineer (a family physician trained in or having expertise in medical informatics[18] and database design.) The knowledge teams define the content for the simulation and select key references (with particular attention to existing evidence-based guidelines) and concepts to include in the knowledge base.

Each knowledge team undergoes an initial orientation to the simulation architecture, including introduction to Bayesian belief network concepts and an overview of the simulation methodology. The team then constructs the parallel and belief network structures needed for simulation. Additionally, the knowledge teams will periodically review knowledge content for updates and revisions.

The development team has accomplished extensive knowledge instantiation for hypothyroidism, essential hypertension, otitis media, duodenal ulcer disease, major depression, reactive airways disease, and abnormal Papanicolaou smears. Comorbidities, such as lipid disorders, renovascular disease, and allergic disorders, have also been added for support of the primary domain simulations. The Board ultimately intends to create simulations for 25 to 30 disease states; external consultants have advised the development team that this number will provide sufficient breadth to support certification decisions (J. Veloski, personal communication, 1995.)

The Simulator

The simulator consists of a computer program that uses knowledge base information to create credible patient cases. The testing paradigm consists of physician queries and interventions and simulator responses. These cases can evolve as a result of elapsed time or in response to candidate interventions. Bayesian belief networks[19] and supporting knowledge base structures interpret, integrate, and update information about a particular patient. As the candidate conducts queries about patient symptoms, physical findings, laboratory results, and clinical test results, the simulator uses belief networks to formulate context-appropriate responses. Similarly, when the candidate orders a particular treatment or therapy, belief networks update the patient's current status. The simulator program also keeps track of time, and the patient evolves as time

passes and in response to the candidate's therapeutic interventions.

The Interface

The early development efforts coincided with the dominance of the Microsoft DOS operating system and predated the advent of the World Wide Web. The evolution of the Windows operating system, improvements in computer hardware, and availability of de facto standard Internet browser platforms coincided with maturation of the simulation system to enable delivery of patient simulations over the Internet. The World Wide Web provides a means for convenient access to the system for candidate practice and will facilitate the candidate registration process. Standard Web browsers enable access to the system without requiring the installation of dedicated, complex computer software packages. This accessibility will allow candidates to become well acquainted with system function before their certification and recertification examinations.

The system uses the Microsoft Internet Explorer Web browser to provide a familiar-appearing interface for the simulations. The development team has worked with Medical Learning Company (Wellesley, Mass) to create the current interface (the system was demonstrated at the American Academy of Family Physicians Scientific Assembly in San Diego, October, 2002.) The interface presents components found in electronic medical records, but it cannot currently be viewed as a model for such systems. As yet no broadly accepted standards exist for electronic medical records, but we anticipate that such guidance will become available in the near future.[20] As standards develop, the interface will evolve to incorporate accepted standard electronic medical record characteristics.

The Administrative Database

The system includes an administrative database to maintain candidate registration information and eligibility for specific examination components. Additionally, this database keeps a record of candidate actions during simulations, thus providing information needed for scoring and for responding to candidate inquiries regarding assessed scores.

Traditional Item Presentation

Family physicians diagnose and manage multiple disease entities[21]; therefore, the computer-based

Table 1. Number of Team Members Responding to Opinions Regarding Their Participation in Knowledge Development Activities: Results of 7-Point Likert Scale (1 = agree, 7 = disagree) Survey of Knowledge.

Opinions	Agree			Neutral			Disagree	Average Score
	1	2	3	4	5	6	7	
I feel comfortable with the knowledge development process	4	11	3	0	0	0	0	1.94
I understand the concept of belief networks	2	9	5	2	0	0	0	2.39
I understand the concept of parallel networks*	6	10	1	1	0	0	0	1.83
Orientation was sufficient to allow the team to proceed with knowledge development	12	6	0	0	0	0	0	1.33

*Parallel network concepts described in Sumner et al.[12]

examination will also include multiple-choice questions similar to those used on the current paper-based test to address less common but important clinical entities for which simulations will not be developed.

The ABFP has developed a multiple-choice examination system that has been used successfully by the American Board of Radiology, the American Board of Pathology, and the American Board of Psychiatry and Neurology for their certification, recertification, and subspecialty examinations (Richard Rovinelli, personal communication, 2002.) The ABFP also used this system to administer the examination for the Certificate of Added Qualification in Sports Medicine in 2002.

The item format mirrors the paper-based examination closely. The questions contain a stem, a question, and possible answers. "Radio buttons" lie adjacent to each question, and the candidate selects the correct answer by clicking the corresponding button with the pointing device (mouse).

Scoring

The dynamic simulation technique poses new challenges for scoring candidate performance. Examinations can be structured so that all candidates start out with identical case presentations. In this context, traditional psychometric models can be used to assess initial candidate actions.[22] As the cases evolve, however, individual candidates will experience unique content, and traditional scoring theory might not apply in these circumstances.[23]

The ABFP has explored several possible alternative approaches. The National Board of Medical Examiners uses expert judgment of possible candidate actions to score USMLE Step 3 simulations.[24] This technique has worked for the National Board

of Medical Examiners because their simulations represent authored cases that will yield to psychometric models, such as item-response theory.[24] The ABFP simulator produces a potentially infinite universe of unique simulation scenarios for which traditional scoring methods do not appear to apply.[25] The development team has therefore explored alternative scoring approaches. The most promising techniques derive from information theory[26] and decision theory.[25–27] The ABFP and the development team are currently conducting research building on analytical constructs described by Downs and Friedman.[25,27,28]

Early Results

The work to date has focused on developing and perfecting the simulation algorithms and populating the knowledge base. Knowledge team development efforts have shown that practicing family physicians, irrespective of their familiarity with computer technology, grasp quickly the supporting concepts (such as Bayesian belief network formulations) and structures needed in the knowledge base. Table 1 summarizes knowledge team members' responses to queries regarding their participation in the knowledge development process.

A prototype presentation system was submitted to field-testing at a number of residency programs in 2001. Feedback from that experience drove the current Web browser appearance and implementation (internal evaluation documents, Assessment Technologies, Inc, June 2001.)

The ABFP began field-testing the current system in early 2003 and anticipates including simulations in the certification and recertification examinations in late 2003 or 2004. The simulations will ultimately encompass 25 to 30 content areas com-

monly managed by family physicians. Functioning simulations currently exist for hypertension, hypothyroidism, otitis media, depression, and duodenal ulcer. Knowledge base development for reactive airway disease and abnormal Papanicolaou smears is nearing completion, and work will begin in 2003 on chronic heart failure and ischemic heart disease.

This system represents a substantial change from the traditional ABFP examination format. To become familiar with the system, candidates will have access to the simulator on a secure Web site via the Internet for practice purposes. The actual certification and recertification examinations will occur initially at testing centers throughout the country. Local intranet networks at each center will display traditional multiple-choice test items and patient simulations. This approach enables, indeed necessitates, more flexible timing for individual candidates' testing sessions. Testing center sites will have limited seating; therefore, the examinations will be administered on multiple dates, providing scheduling options that candidates do not have with the current once-yearly testing format.

Ultimately, a Web-based presentation will allow ABFP certification and recertification candidates to access the process from home, office, or other venues at times convenient to them. This accessibility promises to limit the disruption and cost to candidates' practices that result from the current use of a limited number of geographically distinct test centers. Moving to a widely distributed process will entail security issues that ABFP has not encountered previously. For example, how will the Board verify that the person who logs in is actually the person sitting at the keyboard? Passwords and biometric methods, such as fingerprint recognition and retinal scanning, can verify the identity of the person accessing the system (a good review of these methods appears at http://www.computer.org/itpro/homepage/jan_feb01/security3.htm) but cannot assure that the same person actually remains at the keyboard to complete the activity. Until this issue can be resolved satisfactorily, the Board will continue to deliver the examination through multiple regional testing centers, but with greater frequency and flexibility than possible with the current paper-based test.

Summary
The family practice certification and recertification examinations have relied on paper-based objective test formats since the inception of the ABFP. This approach produces statistically reliable and reproducible results but unfortunately provides only limited ability to assess a candidate's abilities beyond cognitive skill. Simulation provides an additional potential probe into decision making and patient management abilities. Whether this method will yield valid and reproducible information distinct from that gained in paper-based examinations represents an as yet unproved hypothesis.[25] The ABFP recognizes the need to conduct extensive field-testing of the simulations and scoring methodologies before using this technology for making high-stakes certification decisions. Nevertheless, computer-administered examinations will promote testing flexibility not available with the current paper-based testing process. Furthermore, although the simulation architecture requires substantial knowledge acquisition and maintenance, and scoring challenges necessitate nontraditional measurement approaches, moving the examination to an electronic format represents a major step toward making certification and recertification a more flexible, prospective, ongoing process for family physicians.

References
1. Epstein R, Hundert E. Defining and assessing professional competence. JAMA 2002;287:226–35.
2. Bowles L. The challenge and the opportunity. In: Mancall E, Bashook P, editors. Recertification: new evaluation methods and strategies. Evanston, Ill: American Board of Medical Specialties, 1994:3–12.
3. Barnett G. The use of a computer-based system to teach clinical problem solving. Comput Biomed Res 1974;4:301–19.
4. Barnett G, Hoffer EP, Famiglieti K. Computers in medical education: present and future. Proceedings of the Seventh Annual Symposium on Computer Applications in Medical Care. Washington, DC: IEEE Press, 1983:11–13.
5. Harless W, Farr N, Zier M, et al. MERIT - an application of CASE. In: Deland E, editor. Information technology in health science education. New York: Plenum Press, 1978:565–9.
6. Reshetar R. An adaptive testing simulation for a certifying examination. Paper presented at the Annual Meeting of the American Educational Research Association, April 1992, San Francisco.
7. Friedman R. A computer program for simulating the patient-physician encounter. J Med Educ 1973;48:92–7.
8. Clyman S, Orr NA. Status report on the NBME's computer-based testing. Acad Med 1990;65:235–41.

9. Solomon DJ, Osuch JR, Anderson K, et al. A pilot study of the relationship between experts' ratings and scores generated by the NBME's computer-based examination system. Acad Med 1992;67:130–2.

10. Dev P, Hoffer EP, Barnett GO. Computers in medical education. In: Shortliffe EH, Perreault LE, editors. Medical informatics: computer applications in health care and biomedicine. New York: Springer, 2001:610–37.

11. Sumner W 2nd, Truszczynski M, Marek VW. A formal model of family medicine. J Am Board Fam Pract 1996;9:41–52.

12. Sumner W 2nd, Truszczynski M, Marek V. Simulating patients with parallel health state networks. Proc AMIA Symp 1998:438–42.

13. Sumner W 2nd, Marek VW, Truszczynski M. Designing a knowledge base to support family practice certification examinations. Proc AMIA Symp 1993:909.

14. Sumner W 2nd, Truszczynski M, Marek VW. Data transformations for patient simulations. Proc AMIA Symp 1995:970.

15. Sumner W 2nd, Hagen M. Using Bayesian networks as a scripting language. Proc AMIA Symp 1999:1171.

16. Sumner W 2nd, Hagen MD, Rovinelli R. The item generation methodology of an empiric simulation project. Adv Health Sci Educ 1999;4:49–66.

17. Duisterhout JS, van Mulligen EM, van Bemmel JH, van der Maas AAF. Database management. In: van Bemmel JH, Musen MA, editors. Handbook of medical informatics. AW Houten, Netherlands: Bohn Stafleu Van Loghum, 1997:60–6.

18. What is Medical Informatics? In: van Bemmel JH, Musen MA, editors. Handbook of Medical Informatics. AW Houten, Netherlands: Bohn Stafleu Van Loghum, 1997:XXXI-XXXIV.

19. Jensen FV. Causal and Bayesian networks. In: Jensen FV, editor. Bayesian networks and decision graphs. New York: Springer, 2001:3–34.

20. President's Commission on the Electronic Health Record. American Medical Informatics Association [Website]. Available at: www.amia.org. Accessed 10 May 2002.

21. Rosenblatt RA, Hart LG, Gamliel S, Goldstein B, McClendon BJ. Identifying primary care disciplines by analyzing the diagnostic content of ambulatory care. J Am Board Fam Pract 1995;8:34–45.

22. Norcini J. Psychometric issues in computer-based testing. In: Mancall EL, Bashook PG, Dockery JL, editors. Computer-based examinations for board certification. Evanston, Ill: American Board of Medical Specialties, 1996:53–61.

23. Emberton S. Generating items during testing: psychometric issues and models. Psychometrika 1999;64:407–33.

24. Melnick D. The experience of the National Board of Medical Examiners. In: Mancall EL, Bashook PG, Dockery JL, editors. Computer-based examinations for board certification. Evanston, Ill: American Board of Medical Specialties, 1996:111–20.

25. Friedman C, Downs S. Psychometric aspects of computer-based testing. Alternatives to current methods: decision-theoretic approaches. In: Mancall EL, Bashook PG, Dockery JL, editors. Computer-based examinations for board certification. Evanston, Ill: American Board of Medical Specialties, 1996:63–74.

26. Krippendorff K. Information theory: structural models for qualitative data. Newbury Park, Calif: Sage, 1986.

27. Downs SM, Friedman CP, Marasigan F, Gartner G. A decision analytic method for scoring performance on computer-based patient simulations. Proc AMIA Annu Fall Symp 1997:667–71.

28. Downs SM, Marasigan F, Abraham V, Wildemuth B, Friedman CP. Scoring performance on computer-based patient simulations: beyond value of information. Proc AMIA Annu Fall Symp 1999:520–4.