

Simulations in the United States Medical Licensing ExaminationTM (USMLETM)

G F Dillon, J R Boulet, R E Hawkins, D B Swanson

Qual Saf Health Care 2004;13(Suppl 1):i41–i45. doi: 10.1136/qshc.2004.010025

Over the last several years there has been much attention focused on the detection and remediation of problems that pose potential threats to patient safety and that interfere with the provision of effective care. It has been noted that changes in medical education and assessment are integral to eventual improvement in this area. Within the assessment system used to licence physicians in the United States, there has been an evolution of assessment formats intended to improve the measurement of knowledge and skills, including the recent development of computer based patient simulations and clinical skills assessments. A number of new testing formats intended to further enhance assessment of critical knowledge and skills will be available in the near future.

From a test design perspective, some of the IOM recommendations present implementation challenges. For example, the IOM recommended focus on medical teams seems out of place in the context of a medical licensing system aimed at individual clinicians. The primary focus of the medical licensing examination is on an individual's proficiency at playing a defined role on that team. Nevertheless, the idea of "team" can become an important aspect of the medical context presented in test material, especially in the newer performance assessments that have become an important component in modern testing. Such assessment formats can create a sense of team through carefully scripted scenarios in the unfolding simulations and through the requirements of effective communication of findings after clinical encounters.

This paper is intended to provide a review of the examination system currently used as part of the requirements for a medical licence in the United States and, within this context, to describe attempts to assess the knowledge and skills that might reduce the likelihood of problems in healthcare delivery. Relatively new assessment formats are reviewed, including those intended to simulate a real physician-patient encounter. The evolution and future direction of such formats are discussed.

Although the specific context and processes will vary, there are a number of organisations and groups around the world that have a similar responsibility: to determine whether physicians have the skills and knowledge requisite for the safe and effective care of patients. This determination is typically aided by the results of an assessment programme, the extent and complexity of which will vary depending upon the nature of the knowledge and skills to be assessed, upon the psychometric challenges that such assessments bring, or upon a variety of economic, political, legal, ethical, and other issues. To the benefit of these efforts, advances in technology and measurement science have allowed more precise measurement of areas already addressed, and have permitted the assessment of areas not previously addressed because of practical or psychometric difficulties. These advances are particularly useful as organisations involved in such assessments turn their attention to issues identified by the medical community as critical to patient safety. Examples of such issues can be seen in the recent series of reports by the Institutes of Medicine (IOM), which have highlighted fundamental problems in the healthcare system. These reports suggest that the cited problems threaten the quality of patient care in the United States.^{1,2} Recognising the impact that licensing examinations have on influencing perceptions about the critical skills and knowledge that are necessary to combat threats to the quality of healthcare, the IOM reports also include specific recommendations for the licensing examination system.³

THE US MEDICAL LICENSING SYSTEM

A licence to practice medicine in the United States is granted by the licensing authority for the state in which the individual intends to practice. The licence is for the general practice of medicine and does not restrict the individual physician in terms of area of specialisation. In the process of granting the initial medical licence the state requires that the individual meet certain educational and assessment requirements. For medical students trained in the United States or Canada, the educational requirement is met by graduation from an accredited medical school. Students trained outside the United States or Canada must obtain certification by the Educational Commission for Foreign Medical Graduates (ECFMG) which, in part, requires documentation of undergraduate medical training.

Abbreviations: CCS, computer based case simulation; CK, clinical knowledge; CS, clinical skills; ECFMG, Educational Commission for Foreign Medical Graduates; IOM, Institutes of Medicine; MCC, Medical Council of Canada; MCQs, multiple choice questions; NBME, National Board of Medical Examiners; PMPs, patient management problems; SPs, standardised patients; USMLE, United States Medical Licensing Examination

See end of article for authors' affiliations

Correspondence to: Gerard F Dillon, PhD, 3750 Market St, Philadelphia, PA, 19104, USA; gdillon@nbme.org

Key messages

1. Physician assessment is a critical feature in the process intended to determine readiness for independent practice.
2. Recent reports on medical errors recognise the role that assessing physician readiness can have in reinforcing the skills and knowledge important to reducing these problems.
3. Testing formats currently exist that address a variety of skills and knowledge critical to the safe and effective practice of medicine.
4. The challenges to developing reasonable and useful testing formats, though considerable, continue to be successfully addressed by the measurement community.
5. Advances in measurement technology and science can and will allow the measurement of important skills and knowledge that were not previously addressed.

For students receiving the MD degree from a US or Canadian medical school and for all students trained outside of the United States, the assessment requirement for a medical licence is currently met by successful completion of the United States Medical Licensing Examination (USMLE). The USMLE programme is jointly sponsored by the National Board of Medical Examiners (NBME) and by the Federation of State Medical Boards (FSMB). It was first administered in 1992, representing, at that time, a replacement for the multiple examination pathways previously available to licence allopathic physicians.

USMLE is intended to assess a physician's ability to apply knowledge, concepts, and principles, and to demonstrate fundamental patient centred skills, that are important in health and disease and that constitute the basis of safe and effective patient care. There are three components, or "Steps," in the USMLE examination programme.⁴

- Step 1 focuses on the concepts of science basic to the practice of medicine, with special emphasis on principles and mechanisms underlying health, disease, and modes of therapy. Step 1 is a one day, computer delivered examination, made up of multiple choice questions (MCQs).
- Step 2 has two components. The first, known as the clinical knowledge (CK) examination, is a one day, computerised MCQ test intended to assess whether the individual possesses the medical knowledge and understanding of clinical science considered essential for the provision of patient care under supervision. The second component of Step 2 is the clinical skills (CS) component. Step 2 CS is a standardised, patient based examination intended to directly assess the examinee's data gathering and communications skills. Step 2 CS is described in more detail in a subsequent section.
- Step 3 is a two day examination combining MCQs and computer based case simulations (CCSs). It is intended to assess whether the individual can apply medical knowledge and understanding of biomedical and clinical science essential for the unsupervised practice of medicine. The CCS format will also be the focus of discussion in a later section.

In the short history of USMLE, and in the longer history of testing in the medical licensing system generally, there have been numerous attempts to develop assessment formats that provide more authentic representations of patient care. These efforts, in part, reflect dissatisfaction with assessing

physician readiness solely based upon performance in MCQ examinations, but the new formats are not without their own set of challenges.

HISTORY OF SIMULATION USE ON MEDICAL LICENSING EXAMINATIONS

In this context, the term "simulation" is being used in a relatively broad sense. It represents any attempt to reproduce, during testing, relevant aspects of the medical practice environment. This reproduction can vary from the replication of a few isolated features of a clinical task to the recreation of the entire clinical context. There are numerous examples of such attempts, going back to the beginnings of the examination systems used to support medical licensure.

The first NBME examinations, given in 1916, lasted several days and incorporated essay, laboratory, oral, practical, and bedside components.⁵ As an example, the practical examination in surgery required examinees to suture together two segments of a dog intestine. Performance was assessed by determining if the sutures could withstand a prescribed level of water pressure.

In 1922, the NBME's examination programme was restructured. The first two components required for NBME certification assessed understanding of the basic biomedical sciences and the fundamentals of clinical medicine, both through essay questions. The final component included observed patient encounters followed by oral examinations regarding those encounters. This structure persisted until the late 1950s, when studies of the bedside oral exam clearly documented its psychometric inadequacies. Scores were found to provide more information about the examiner than about the examinee. Agreement between examiners observing examinees with different patients was at near-chance levels,⁶ reflecting both differences in examiner stringency and variation in the quality of an examinee's performance from one clinical situation to another. High costs and logistical difficulties in administering this type of examination for the rapidly growing post-war cohort of medical students also contributed to its demise.⁵ Over the next two decades, test developers experimented with a number of assessment methods, all intended to recover the assessment of clinical skills that was lost with the elimination of observed clinical encounters. Initially, motion pictures of clinical encounters were projected to examinees, who answered MCQs about the portrayed encounter. The key problem with this format was logistical: it was difficult to standardise the test administration conditions for projection of films.³ During the late 1960s, in an effort to pose more realistic challenges to medical decision making skills, patient management problems (PMPs) were introduced. These multi-step problems began with an "opening scenario" that provided a brief description of a patient care situation. The examinee then proceeded through a series of "scenes" in which additional information was gathered (history taking, physical examination, and laboratory scenes), followed by one or more scenes in which patient management activities were initiated. "Latent image" pens were used to select actions and reveal feedback summarising the consequences of the actions. PMPs were commonly used on US and Canadian medical licensing and certification examinations until the late 1980s, when the following problems led to their elimination.⁷ Like bedside oral examinations, relatively small numbers of PMPs were included on the examination, and the small sample of cases resulted in relatively unreliable scores. Additionally, in many clinical situations, a broad range of patient management strategies are possible, and it was sometimes difficult to develop scoring keys that appropriately rewarded alternate strategies that were similar in quality. Variation in examinees' response style was also problematic; scoring keys

tended to reward the thorough examinee and penalise the efficient examinee, leading to scores that, in part, reflected examinees' propensity to take action in an uncertain situation, rather than just their clinical decision making skills.⁸

CURRENT USE OF SIMULATIONS ON USMLE

Simulations vary considerably in fidelity: one could view MCQs as simulations (at least if they begin with a patient description and require examinees to indicate a clinical decision) at the low end of the fidelity continuum, providing an assessment of examinees' proficiency in applying their knowledge to descriptions of case situations. Assessments using standardised patients (SPs) lie at the other end of the fidelity continuum, providing a realistic context for measuring the skills involved in taking a history and performing a physical examination. Computer based case simulations fall in between. This section provides an overview of how each is currently used in USMLE.

Patient based MCQs

Patient based MCQs have been used for decades on licensing examinations; they have appeared in all three steps since the introduction of USMLE. On Step 1, these take the form of brief descriptions of patient care situations, followed by questions challenging examinees to use their understanding of basic biomedical science to explain or predict patient findings.⁹ Roughly 60% of the items on Step 1 currently take this form. In Step 2, virtually all items begin with a description of a clinical situation; the patient presentations are longer and less prototypic than the patient descriptions on Step 1. Examinees must differentiate important from incidental findings and indicate a clinical decision, generally a diagnosis or the next step in patient care.¹⁰ Step 3 MCQs also provide detailed descriptions of physician-patient encounters. Items are organised by the encounter setting (for example, office or emergency department) and are often presented in sets termed "case clusters" in which a series of MCQs address different facets of an unfolding clinical situation.¹¹

From a content sampling perspective, because relatively large numbers of MCQs can be administered per hour of testing time, MCQs on all three USMLE steps provide an efficient method for assessment of decision making skills. The degree of fidelity depends, in part, upon the length of the patient description, the level of detail provided, and the extent to which patient findings are provided in an interpreted versus an undigested format.¹² In the near future, it is likely that the fidelity of MCQs on USMLE will increase, as all three steps take advantage of computer based test administration to incorporate multimedia into item "stems," thus enriching patient presentations.

Computer based case simulations

Three decades of research effort eventually culminated in the inclusion of uncued, interactive CCS on USMLE Step 3 in 1999.¹³ In CCS, the examinee is presented with a brief description of a patient, including a chief complaint and a brief history.^{11 14} From that point forward, the case unfolds as the examinee works up and manages the computer simulated patient, obtaining diagnostic information, ordering therapeutic interventions, and monitoring patient progress. Any of several thousand diagnostic and therapeutic manoeuvres can be requested by the examinee in free text on an "order sheet." As simulated time passes, the patient's condition changes based on the underlying medical problem and the examinee's interventions; results of tests are reported and the impact of interventions must be monitored. Examinees are scored on CCS using an algorithm that essentially compares their patient management strategies with policies obtained

from experienced clinicians.¹⁵ Examinees must balance thoroughness, efficiency, timeliness, and avoidance of risk in responding to clinical situations, with dangerous and unnecessary actions lowering scores. Although CCS cases have proven expensive to develop, administer, and score, psychometric analyses have indicated that CCS cases measure some of the management knowledge and skills that are not easily addressed by MCQs, and they do so with a reasonable degree of precision.^{16 17}

Clinical skills assessment using standardised patients

In mid 2004, a standardised patient based clinical skills examination was added as a component of Step 2 of USMLE. Developed in collaboration with the ECFMG the purpose of this component, called Step 2 CS, is to ensure the public that successful candidates for licensure are competent in the fundamental clinical skills required for safe and effective patient care. These clinical skills include taking a relevant medical history, performing an appropriate physical examination, communicating effectively with the patient, clearly and accurately documenting the findings and diagnostic hypotheses from the clinical encounter, and listing appropriate initial diagnostic studies.

The examination consists of 12 encounters with SPs portraying common and important clinical problems. History taking questions and physical examination manoeuvres are recorded by SPs using case specific dichotomous checklists that are completed after each encounter. Communication and interpersonal skills, and spoken English proficiency, are evaluated by SPs using generic rating scales. The patient note, a clinical summary recorded after the encounter, is scored by physician raters using holistic methods. Examinees are assessed on three subcomponents: Integrated Clinical Encounter (ICE), which includes data gathering (history taking and physical examination) and the patient note; Communication and Interpersonal Skills (CIS); and Spoken English Proficiency (SEP). Examinees must pass all three subcomponents in order to pass Step 2 CS overall.

Students graduating from MD granting US and Canadian medical schools in 2005 or later are required to take USMLE Step 2 CS as part of the requirements for licensure in all United States jurisdictions. USMLE Step 2 CS replaces the Clinical Skills Assessment (CSA[®]) of the ECFMG, as an essential requirement for ECFMG certification and licensure for graduates of medical schools located outside of the US and Canada, who are seeking postgraduate training opportunities in the United States.

Implementation of USMLE Step 2 CS follows the introduction of similar SP based examinations for high stakes decision making in other countries. The Medical Council of Canada (MCC) has included a multi-station SP based examination as part of the MCC Qualifying Examination for licensure since 1993.^{18 19} The General Medical Council of the United Kingdom has included a SP based assessment as one component of the Professional Linguistics and Assessment Board (PLAB) examination since 1998.²⁰ The above mentioned ECFMG CSA also has been in place since 1998.²¹ The introduction of SP based examinations for licensure stems from two important and converging concepts: firstly, a growing body of evidence and experience supporting the value of SP based methods for assessing clinical skills, and secondly, a re-emerging acknowledgment that these patient centred skills are critical to safe and effective patient care.

High stakes clinical skills examinations using SPs are developed and administered in a manner that is based upon and supported by approximately four decades of research and experience. Indeed, a significant part of the research into this particular form of simulation was conducted by the

organisations (NBME and ECFMG) collaborating in USMLE Step 2 CS development. Similar to other forms of simulation, the use of SPs for assessment offers clear benefits, including more certain availability of relevant test material during examinations, absent risk to real patients, an environment and set of examinee tasks that appear more authentic relative to the attributes assessed, and decreased cost compared to the use of clinician raters.

Numerous studies suggest that the use of SPs allows for the valid assessment of basic clinic skills.^{22–26} Careful attention to SP recruiting and training approaches and the use of process checklists and rating scales for scoring examinee performance enhance the reliability of such assessments.^{22 27–29} As examinee performance varies substantially across clinical content areas, the inclusion of a broad sample of cases is necessary to optimise reliable assessment of clinical skills.^{22 30} For this reason examinee–SP encounters are typically arranged in a series of stations; examinees are required to demonstrate one or more of the skills of interest in response to varying patient presentations.

It is clear that the clinical skills assessed by USMLE Step 2 CS are essential to safe and effective patient care. The medical history and physical examination contribute important information to patient diagnosis and management.^{31 32} Adequate communication and interpersonal skills are associated with enhanced patient satisfaction and improved clinical outcomes, as well as decreased risk for malpractice litigation.^{33–36} Poor written documentation in the medical record has been identified in ambulatory and inpatient settings as having potential implications for healthcare quality.^{37–39} Enhancements planned or under consideration for inclusion in Step 2 CS will allow for an even more robust assessment of physical examination skills, without posing safety risks to SPs.

The USMLE, in serving the licensure process, targets assessment at the individual examinee, assuring the public that successful candidates, regardless of education and experience, have met a national standard for performance in those domains assessed by its three Steps. The inclusion of simulation methods, SPs in Step 2, and CCS in Step 3, allows for a richer assessment of patient centred skills and management approaches that are critical to patient care. To the degree that USMLE can determine that a physician-to-be may lack these skills at a very fundamental level, the potential impact for patient safety seems apparent. While the USMLE does not measure the performance of teams per se, attributes that allow examinees to function effectively on a healthcare team are included in the examination. For example, adequate verbal and written communication skills, assessed in Step 2 CS, are essential components of effective teamwork (although the extent to which communication skills exercised within the doctor–patient encounter generalise to effective communication within healthcare teams remains to be established). And of course the inclusion of a patient centred component in the licensure examination underscores the purpose for which healthcare teams are constructed—to provide beneficial care for patients.

FUTURE OF SIMULATIONS FOR MEDICAL LICENSING

It is likely that technological and educational advances will lead to the enhancement of testing formats and scoring modalities already used in USMLE. For example, many medical schools have incorporated computer based training modules, part task trainers that focus on specific procedures and body parts (for example, breast models, pelvic models) and full scale integrated simulators for training purposes.^{40 41} Despite the fact that their use has been primarily for summative assessment purposes and has been somewhat limited due to cost and the lack of efficient, reliable, and valid

scoring rubrics, these formats will likely play a role in examination systems for licensure. Like the CCS component of Step 3 and the CS part of Step 2, once there is evidence to support the use of scores from these formats, these new simulation modalities will certainly be embraced.

While there are a number of new simulation modalities that may be applicable for high stakes licensing examinations, the most likely developments in the near term will relate to the further refinement of current methods, including scoring systems. For CCS, a tremendous amount of research was completed to develop and validate the scoring systems.^{17 42–46} Even so, new models (for example, neural networks) may eventually lead to even more reproducible scores and more valid assessment decisions. For Step 2 CS (standardised patients), the introduction of different types of cases (for example, telephone consultation, focused counselling) could enhance the content validity of the assessment. In addition, provided that the psychometric adequacy of the scores is shown to be sufficient, future cases could be constructed so that team dynamics and communication could be assessed. Currently, doctor–patient communication skills are evaluated as part of Step 2 CS; an extension to doctor–nurse communication, or other relevant teams, although complicated, is certainly possible. Part task trainers would also seem to be an appropriate add on. Here, select physical examination skills (for example, breast examination, pelvic examination) that are difficult to measure in SP based assessments could be incorporated in some of the stations.^{47 48} This addition to Step 2 CS would augment the assessment of technical proficiency in physical examination performance, allowing examinees the opportunity to identify important abnormalities. In terms of scoring simulation based assessments, there are definitely some enhancements and modifications that could be considered. For example, the approach to statistically modelling expert scorers, used for CCS, or some variant, may be applicable to score the written clinical summaries (patient notes) that are produced by examinees following each of their encounters with the SP. These notes, by design, are intended to measure written communication with the healthcare team, and are an integral part of Step 2 CS.⁴⁹

There are also a number of new technologies that could eventually find their way into high stakes medical licensure examinations. Virtual reality and haptic feedback trainers have been used to train physicians in areas such as minimally invasive surgical procedures and vascular interventions.⁵⁰ Unfortunately, the cost for these systems can be high, especially if they are to be used for large scale assessments where many thousands of examinees must be tested. More importantly, these technologies generally target very specialised skills. From a certification or licensure perspective, where the focus centres on measuring fundamental skills as reliably as possible, it may be inefficient and impractical to use these systems. Additional research will be necessary before these types of trainers can be incorporated in high stakes summative assessments.

Life size mannequins (integrated simulators) with realistic airway and cardiovascular attributes have been used to train physicians and other healthcare professionals. For medical licensing examinations, these simulators suffer from some of the same drawbacks as virtual reality and haptic feedback trainers. They are costly and generally target more specialised skills. Moreover, while a number of scoring systems have been developed,^{51 52} they have yet to undergo the scientific scrutiny that has taken place for SP assessments.⁵³ Nevertheless, as the cost of these mannequins declines, and additional psychometric studies are completed, they could have a unique role within the licensure process, especially for higher order skills. Firstly, mannequins can be used to model

rare events, especially those where medical errors would not be reversible if a “real” patient were being managed. This is important in that some clinical skills are difficult, if not impossible, to measure, even with well trained SPs. Secondly, since real time responses to therapeutic interventions can be modelled, the management of patient conditions such as drug interventions can be assessed. Thirdly, it is possible to develop scoring systems that are based on measurable patient outcomes, something that can be difficult to do, at least in a consistent way, with SP based assessments. Finally, for healthcare teams (for example, trauma) it is possible to assess joint patient care efforts, including multidisciplinary communication skills, in a standardised environment. If test content considerations, logistics, and scoring issues can be addressed, the use of integrated simulators as part of licensure examinations may be forthcoming.

Authors' affiliations

G F Dillon, R E Hawkins, D B Swanson, National Board of Medical Examiners, Philadelphia, Pennsylvania, USA

J R Boulet, Educational Commission for Foreign Medical Graduates, Philadelphia, Pennsylvania, USA

REFERENCES

- Kohn LT, Corrigan JM, Donaldson MS, eds. *To err is human: building a safer health system*. Washington DC: Institute of Medicine, National Academies Press, 1999.
- Institute of Medicine. *Crossing the quality chasm: a new health system for the 21st century*. Washington DC: Institute of Medicine, National Academies Press, 2001.
- Greiner AC, Knebel E, eds. *Health professions education: a bridge to quality*. Washington DC: Institute of Medicine, National Academies Press, 2003.
- Federation of State Medical Boards Inc, National Board of Medical Examiners. *USMLE 2004 bulletin of information*. Philadelphia: FSMB and NBME, 2003.
- Hubbard J, Levit E. *The National Board of Medical Examiners: the first seventy years*. Philadelphia: National Board of Medical Examiners, 1985.
- Hubbard JP. *Measuring medical education*. Philadelphia: Lea and Febiger, 1971.
- Swanson D, Norcini J, Grosso L. Assessment of clinical competence: written and computer-based simulations. *Assessment and Evaluation in Higher Education* 1987;12:220–46.
- Swanson DB, Norman GR, Linn RL. Performance-based assessment: lessons from the health professions. *Educational Researcher* 1995;24:5–11.
- Federation of State Medical Boards Inc, National Board of Medical Examiners. *2004 USMLE Step 1 content description and sample test materials*. Philadelphia: FSMB and NBME, 2003.
- Federation of State Medical Boards Inc, National Board of Medical Examiners. *2004 USMLE Step 2 content description and sample test materials*. Philadelphia: FSMB and NBME, 2003.
- Federation of State Medical Boards Inc, National Board of Medical Examiners. *2004 USMLE Step 3 content description and sample test materials*. Philadelphia: FSMB and NBME, 2003.
- Swanson D, Case S. Trends in written assessment: a strangely biased perspective. In: Harden R, Hart I, Mulholland H, eds. *Approaches to the assessment of clinical competence: part 1*. Norwich, England: Page Brothers, 1992:38–53.
- Melnick DE, Dillon GF, Swanson DB. Medical licensing examinations in the United States. *J Dent Educ* 2002;66:595–9.
- Clyman S, Melnick D, Clauser B. Computer-based case simulations from medicine: assessing skills in patient management. In: Tekian A, McGuire C, McGaghie W, eds. *Innovative simulations for assessing professional competence*. Chicago: University of Illinois, Department of Medical Education, 1999:29–41.
- Clauser BE, Subhiyah TE, Piemme LG, et al. Using clinician ratings to model score weights for a computer-based clinical—simulation examination. *Acad Med* 1993;68(Suppl):S64–6.
- Clauser BE, Margolis MJ, Swanson DB. An examination of the contribution of computer-based case simulations to the USMLE Step 3 examination. *Acad Med* 2002;77(Suppl):S80–2.
- Dillon GF, Clyman SG, Clauser BE, et al. The introduction of computer-based case simulations into the United States medical licensing examination. *Acad Med* 2002;77(Suppl):S94–6.
- Reznick RK, Blackmore D, Cohen R, et al. An objective structured clinical examination for the licentiate of the Medical Council of Canada: from research to reality. *Acad Med* 1993;68(Suppl):S4–6.
- Reznick RK, Blackmore D, Dauphinee WD, et al. Large-scale high-stakes testing with an OSCE: report from the Medical Council of Canada. *Acad Med* 1996;71(Suppl):S19–21.
- Tomblason P, Fox RA, Dacre JA. Defining the content for the objective structured clinical examination component of the professional and linguistic assessments board examination: development of a blueprint. *Med Educ* 2000;34:566–72.
- Whelan G. High-stakes medical performance testing: the clinical skills assessment program. *JAMA* 2000;283:1748.
- Van der Vleuten CPM, Swanson DB. Assessment of clinical skills with standardized patients: state of the art. *Teach Learn Med* 1990;2:58–76.
- Stillman P, Swanson D, Regan MB, et al. Assessment of clinical skills of residents utilizing standardized patients: a follow-up study and recommendations for application. *Ann Intern Med* 1991;114:393–401.
- Boulet JR, McKinly D, Whelan GP, et al. Clinical skills deficiencies among first-year residents: utility of the ECFMG clinical skills assessment. *Acad Med* 2002;77(Suppl):S33–5.
- Harasym PH, Mohtadi NG, Henningsmoen H. Construct validity of “high stakes” OSCE scores. In: Scherpbier AJA, van der Vleuten C, Rethans JJ, et al. *Advances in medical education*. Kluwer Dordrecht: Academic Publishers, 1997:443–5.
- Vu NV, Marcy ML, Barnhart A, et al. Further evidence of construct validity of standardized patient-based performance examinations. *Teach Learn Med* 1994;6:25–59.
- King AM, Perkowski-Rogers LC, Pohl HS. Planning standardized patient programs: case development, patient training, and costs. *Teach Learn Med* 1994;6:6–14.
- Tamblyn RM, Klass DK, Schnable, et al. Factors associated with the accuracy of standardized patient presentations. *Acad Med* 1990;65(Suppl):S55–6.
- De Champlain AF, Margolis MJ, King A, et al. Standardized patient's accuracy in recording examinees' behaviors using checklists. *Acad Med* 1997;72(Suppl):S85–7.
- Swanson DB, Norcini JJ. Factors influencing reproducibility of tests using standardized patients. *Teach Learn Med* 1989;1:158–66.
- Peterson MC, Holbrook JH, Hales DV, et al. Contributions of the history, physical examination and laboratory investigation in making medical diagnoses. *West J Med* 1992;156:163–5.
- Reilly BM. Physical examination in the care of medical inpatients: an observational study. *Lancet* 2003;362:1100–5.
- Stewart MA. Effective physician-patient communication and health outcomes: a review. *CMAJ* 1995;152:1423–33.
- Suchman A, Roter D, Green M, et al. Physician satisfaction with primary care office visits. *Med Care* 1993;31:1083–92.
- Levinson W, Roter DL, Mullooly JP, et al. Physician-patient communication. The relationship with malpractice claims among primary care physicians and surgeons. *JAMA* 1997;277:553–9.
- Beckman HB, Markakis KM, Suchman AL, et al. The doctor-patient relationship and malpractice. Lessons from plaintiff depositions. *Arch Intern Med* 1994;154:1365–70.
- Beers MH, Munekata M, Storrie M. The accuracy of medication histories in the hospital medical records of elderly patients. *J Am Geriatr Soc* 1990;38:1183–7.
- Cox JL, Zitner D, Courtney KD, et al. Undocumented patient information: an impediment to quality of care. *Am J Med* 2003;114:211–16.
- Soto CM, Kleinman KP, Simon SR. Quality and correlates of medical record documentation in the ambulatory care setting. *BMC Health Serv Res* 2002;2:22.
- Good ML. Patient simulation for training basic and advanced clinical skills. *Med Educ* 2003;37(Suppl):14–21.
- Schuwirth LW, van der Vleuten CP. The use of clinical simulations in assessment. *Med Educ* 2003;37(Suppl):65–71.
- Clauser BE, Subhiyah TE, Piemme LG, et al. Using clinician ratings to model score weights for a computer-based clinical-simulation examination. *Acad Med* 1993;68(Suppl):S64–6.
- Clauser BE, Subhiyah TE, Nungester RJ, et al. Scoring a performance-based assessment by modeling the judgments of experts. *Journal of Educational Measurement* 1995;32:397–415.
- Clauser BE, Margolis MJ, Clyman SG, et al. Development of automated scoring algorithms for complex performance assessments: a comparison of two approaches. *Journal of Educational Measurement* 1997;34:141–61.
- Clauser BE, Ross LP, Fan VY, et al. A comparison of two approaches for modeling expert judgment in scoring a performance assessment of physicians' patient-management skills. *Acad Med* 1998;73(Suppl):S117–19.
- Clauser BE, Swanson DB, Clyman SG. A comparison of the generalizability of scores produced by expert raters and automated scoring systems. *Applied Measurement in Education* 1999;12:281–99.
- Federation of State Medical Boards, National Board of Medical Examiners. *2004 USMLE Step 2 CS content description and general information booklet*. Philadelphia: FSMB and NBME, 2003.
- Pugh CM, Youngblood P. Development and validation of assessment measures for a newly developed physical examination simulator. *Journal of the American Informatics Association* 2003;9:448–60.
- Boulet JR, Rebbecchi TA, Denton EC, et al. Assessing the written communication skills of medical school graduates. *Adv Health Sci Educ Theory Pract* 2004;9:47–60.
- Nackman GB, Bermann M, Hammond J. Effective use of human simulators in surgical education. *J Surg Res* 2003;115:214–18.
- Boulet JR, Murray D, Kras J, et al. Reliability and validity of a simulation-based acute care skills assessment for medical students and residents. *Anesthesiology* 2003;99:1270–80.
- Morgan PJ, Cleave-Hogg D, DeSousa S, et al. High-fidelity patient simulation: validation of performance checklists. *Br J Anaesth* 2004;92:388–92.
- Norcini J, Boulet J. Methodological issues in the use of standardized patients for assessment. *Teach Learn Med* 2003;15:293–7.