

Standardized Patient Encounters

A Method for Teaching and Evaluation

Michael A. Ainsworth, MD; Linda P. Rogers, MS; John F. Markus, MS; Neil K. Dorsey, MD;
Thomas A. Blackwell, MD; Emil R. Petrusa, PhD

The primary goal of medical education is to foster development of clinical competence in trainees at all levels. Variable clinical experience, inconsistent methods of instruction, and ambiguous evaluation criteria undermine this goal. Standardized patients, trained to consistently portray a wide variety of clinical cases, can help overcome many of these educational problems. This article describes the development and application of standardized patients throughout medical training at The University of Texas Medical Branch, Galveston, in the freshman interviewing course, the second-year physical diagnosis course, third-year clerkships, a fourth-year final exercise, and residency training. Development of this program is discussed in the context of a broader literature in medical education, and investigation of variables affecting standardized patient and student performance is reported. Future directions for use of standardized patients in monitoring and promoting the development of clinical competence are discussed.

(*JAMA*. 1991;266:1390-1396)

MEDICAL educators carry dual responsibilities of teaching and evaluation. Traditional methods of medical education place emphasis on transmitting medical knowledge and modeling clinical skills in history taking, physical examination, differential diagnosis, and problem solving. Although most US medical schools share a nucleus of common courses and curricular requirements focusing on medical knowledge, considerable variability exists in clinical instruction.¹ Within clinical clerkships, students at different institutions encounter diverse patient populations in a variety of settings, with a wide spectrum of diseases and presentations.²

From the Department of Internal Medicine (Drs Ainsworth, Dorsey, Blackwell, and Petrusa), and Office of Educational Development (Dr Petrusa, Ms Rogers, and Mr Markus), The University of Texas Medical Branch, Galveston. Dr Petrusa is now at Duke University Medical Center, Durham, NC.

Reprint requests to the Department of Internal Medicine, The University of Texas Medical Branch, UTMB Box 172, Galveston, TX 77550 (Dr Ainsworth).

Even within a single discipline, there may be little consistency or standardization of student experiences.³ The diversity of clinical exposure during medical training suggests further that a graduate's clinical competence also is likely to vary.⁴

Clinical performance includes the skills of medical interviewing, physical examination, differential diagnosis and decision making, and behavioral characteristics expected of all physicians. The purpose of clinical performance assessment is to provide behavior-based evidence that these skills have been mastered. Within the medical education literature, methods for the assessment of clinical performance have been identified as an area in need of immediate improvement.⁵ Future guidelines for accreditation likely will require medical schools to develop a longitudinal database documenting the professional growth of its students, including documentation of clinical skills, further em-

phasizing the critical need for advancement in this area.⁶ Therefore, each institution must be responsible for designing and implementing appropriate methods to verify the competence of its graduates. Following an overview of problems with current methods of developing and evaluating clinical skills, we will describe the historical development and current longitudinal application of standardized patients (SPs) for teaching and evaluating clinical performance of students and residents at The University of Texas Medical Branch (UTMB). Areas of active investigation will be emphasized, followed by a summary of important issues in SP-based assessments and future directions for research.

PROBLEMS WITH CURRENT ASSESSMENT TECHNIQUES

Clinical instruction is incomplete without a means of assessing what has been learned and retained for subsequent use. Traditional assessments of ability, such as multiple choice and other written examinations, primarily emphasize the examinee's capacity for data recall. Although criticized for their misapplication and misinterpretation, such examinations are used widely because they allow query on a broad sampling of knowledge, are easy to administer and score, and produce quantitative data about examinee performance. When administered as standardized examinations to a large pool of students, eg, the certifying examinations of the National Board of Medical Examiners (NBME), potential problems in the reliability of test items can be minimized. Not surprisingly, however, investigators have found inconsistent correlation

between knowledge-based assessments such as examinations administered by the NBME, and clinically based assessments.⁷⁻⁹

Current clinical evaluation methods are also imperfect.^{7,8,10,11} Traditional judgments of clinical competence are based on evaluations by clinical teachers in the setting of patient care. The reliability of this approach is limited by the uncontrolled nature of the clinical environment, especially the lack of standardization among observers, and the infrequent direct sampling of clinical skills. Assessment techniques should measure a student's ability to gather data by interview and physical examination, process this information, and make decisions based on this database. Traditional clinical evaluations often measure the final product of student interaction with patients (differential diagnosis and management), but cast little light on the building blocks of these skills—data acquisition and the process by which clinical problems are solved. One explanation for this shortcoming is the infrequent direct observation of clinical performance. Better methods for assessing clinical skills are clearly needed.

Standardized patients are one resource that has been used extensively for performance-based testing, that is, the direct observation or measurement of performance skills and behaviors required of physicians. Standardized patients are nonphysicians trained to accurately and consistently simulate a patient encounter for teaching or evaluation purposes.⁷ A standardized encounter is constructed by training the "patient" to provide reproducible responses to examinee questions or to simulate consistent physical examination findings. Patients with genuine medical problems, particularly stable physical examination abnormalities, also are valuable as SPs in assessing clinical skills. Through proper training, SPs can improve the objectivity and reliability of the measurement of clinical skills, while minimizing inconsistencies inherent to evaluation based on random clinical encounters, where student knowledge and performance may vary depending on the patient selected.¹²

STANDARDIZED PATIENT PROGRAMS AT UTMB

In this section we will describe the development of an SP program at UTMB and discuss the current application of SPs in the first-year medical interviewing course, second-year physical diagnosis course, third-year clerkships, fourth-year student curriculum, and residency training. Lastly, we

will describe efforts to address unresolved issues of SP teaching and competency assessment.

Standardized patients have been used extensively in the UTMB curriculum since the mid 1970s. The medical school enrolls 200 students each year and has a traditional 4-year curriculum (ie, 2 years of basic science courses, 1 year of clinical clerkships, and 1 year of elective rotations). The SP program at UTMB developed through the efforts of faculty in the university's Office of Educational Development. With the guidance of medical educators aware of the potential value of SP encounters for teaching and with the fiscal and philosophical support of the dean of medicine, UTMB initiated a formal trained-patient program in 1976. This development also occurred at a time of increased emphasis on behavioral sciences in medicine by the NBME, which began including behavioral science topics in their examinations during the early 1970s.

Designed to supplement student experience and allow practice of clinical skills, SP-based encounters are not a substitute for direct contact with patients in the clinical setting. The initial use of SPs was limited to student practice of medical interviewing skills, but over the next decade, use of SP encounters became more sophisticated and widely applied in the curriculum. During the 1980s, the use of SPs as an instructional resource and evaluation method was widely accepted by students and educators at UTMB, and SPs were incorporated into the physical diagnosis course to evaluate students' cumulative interviewing and physical examination skills. In 1983, the Department of Internal Medicine implemented its first objective structured clinical examination (OSCE), a multistation test of specific interviewing and physical examination skills, as an evaluation tool. In 1987, as a consequence of internal curriculum reviews, the faculty of the School of Medicine voted to use SP encounters, with basic science and clinical faculty observers, to ensure faculty involvement in the assessment of the quality of our graduating seniors. In 1990, the Department of Family Medicine added SP encounters to its clerkship and used its faculty to observe and discuss the clinical performance of students.

The broad application of this methodology throughout the curriculum illustrates a major commitment of the university to SP-based assessment. Implementation of such a program and development of new applications would be impossible without an established insti-

tutional program to recruit and train SPs.

Training SPs

The UTMB program currently maintains a pool of approximately 60 SPs (ages 13 to 72 years). These participants are recruited from the community, faculty, staff, and students of UTMB, and from among patients identified by clinical faculty in their group practice. Recruitment, training, and evaluation of SPs is supervised by two full-time staff members. The SPs are often trained for multiple roles (approximately seven roles per year) and are trained according to protocols adapted to course goals and examinees' level of training. Standardized patient performances are observed by staff during their initial portrayal and periodically thereafter to ensure consistency in case presentation. Standardized patients are compensated based on the number of hours and types of cases portrayed, generally \$10 for each hour of training and examination participation. Standardized patients undergoing rectal examination receive greater compensation. Initial training costs, averaging \$20 to \$30 per patient per case, are borne by the institution, while examination and instruction costs are the responsibility of the department administering the SP-based exercise.

Specific case scripts are developed by clinical experts in each area, in collaboration with a patient trainer (educator). The applicable course or school committee then reviews each script to assess the problem's appropriateness for its intended application. For example, an introductory case would involve a chief complaint focused on one organ system with minimal psychosocial, family, or past medical issues. Encounters of greater complexity are constructed for fourth-year students, with a wider differential diagnosis and complicating social or past medical issues. Introductory cases are often modified to make them more diagnostically complex (eg, to encourage students and faculty examiners to consider multiple diagnostic possibilities). Faculty members also conduct physical examinations on the SPs and assess the accuracy and fidelity of the performances each year, with periodic audio- and videotaping of performances to identify SPs who need more frequent monitoring.

The primary objective for SP encounters at UTMB is to assist in instruction of fundamental medical interviewing and physical examination skills. However, as the program has matured, SPs have also been used at more advanced

Table 1.—Use of Standardized Patient Encounters at UTMB*

Course	SP-Based Activity	Duration, min
Introduction to Patient Evaluation	Two medical interviews; basic problems with faculty and SP feedback	30
	Focused examination with SP feedback	30
	Complete examination with SP feedback	60
	Complete interview and examination with SP and faculty feedback	75
Introduction to Clinical Medicine	Final evaluation exercise; complete interview and examination with SP and faculty feedback	90
Internal medicine clerkship	26 focused interviews or examinations	4
Family medicine clerkship	Two complete interviews and examinations with SP and faculty feedback	30
Senior assessment exercise	Complete interview and examination of patient with complex problem; SP and faculty feedback	45
Internal medicine residency	12 focused interviews or examinations with faculty feedback	4

*UTMB indicates University of Texas Medical Branch, and SP, standardized patient.

levels of medical training and have been trained to evaluate and provide feedback to examinees at all levels. Table 1 summarizes the use of SPs in the UTMB undergraduate and graduate curriculum.

Logistical problems with the SP program have been minimized due to the availability of eligible patients in a large university environment, the retention of a stable pool of SPs, and a permanent full-time staff of medical education professionals to ensure uniform training standards and to assist in the development and administration of SP exercises. Difficulties occasionally arise when patients scheduled for specific examinations are unable to participate, or when previously stable physical findings change, necessitating confirmation of a patient's true findings prior to each exercise. While SP turnover is inevitable and anticipated as a result of patient relocation or disease progression and resolution, patient burnout is kept to a minimum by emphasizing to SPs the valuable role they have in medical education. We encourage SPs to limit participation to 1- or 2-hour sessions at a time, provide feedback and retraining, vary SP roles and evaluation tasks, and learn multiple scripts so they may vary their performances.

Introduction to Patient Evaluation: Building History-Taking and Physical Examination Skills

The School of Medicine's Introduction to Patient Evaluation course has included SPs in teaching and evaluating interviewing skills of students for 15 years. In the year-long course, all 200 first-year students meet regularly in small groups in self-directed learning sessions supplemented by faculty facilitators and practice patient interviews in the hospital or clinic. To evaluate medical interviewing under controlled conditions, all students practice these skills

by means of two 30-minute interviews with SPs. These interactions are enriched through direct feedback by SPs to students on data collection, interpersonal skills, and nonverbal communication. Videotapes of these sessions are later reviewed by the groups, with their faculty mentor guiding discussion and providing additional feedback. A range of student cognitive and behavior skills may be challenged by SPs trained to simulate cases of common medical problems confounded by psychological, social, or environmental factors.

In an effort to integrate the interview with the physical examination, this program recently expanded to include practice of elementary physical examination skills. Students now have two 30- to 60-minute sessions with an SP to perform practice examinations. The SPs provide feedback on errors of technique or omission and counsel students on behavioral skills such as appropriate draping of patients and communication skills designed to put patients at ease. As a final evaluation exercise in Introduction to Patient Evaluation, these skills are tested in an integrated fashion with a complete interview and examination of an SP in a 90-minute encounter under direct faculty observation. After feedback and reinforcement from both faculty and SP, students prepare a written account of the patient encounter, including problem list, assessment, and a plan to address the patient's complaint.

Introduction to Clinical Medicine: Integrating Clinical Skills

The Introduction to Clinical Medicine course is a two-phase interdisciplinary program during a medical student's second year of training. Similar in organization, but on a more concentrated scale than Introduction to Patient Evaluation, the initial phase features weekly interaction among student groups and faculty leaders. Consolidation of

physical examination skills is fostered through sequential faculty modeling and student practice of examination techniques, with distinction between normal and pathological findings emphasized through organ-based examination of hospitalized patients.

The second phase integrates this experience through complete interviews and examinations of hospitalized patients, under guidance of the student's faculty mentor. Before entering the clinical years, each student is formally evaluated through a 90-minute SP encounter involving a complete screening history and physical examination. In this summative exercise, the student receives immediate verbal feedback from a highly trained SP on the historical and physical examination items considered appropriate for the encounter. This feedback serves to identify errors and omissions, and provides positive reinforcement to students on the favorable aspects of their performance. The use of SPs in this format has resulted in uniformly positive student response (collected through annual course evaluations) and excellent agreement between SP and faculty ratings of student competence on the exercise.

Clinical Clerkships: Demonstration of Competence

The Department of Internal Medicine at UTMB has capitalized on SP encounters as an evaluation tool for nearly a decade. The most extensive application of this method has been as a summative evaluation of all 200 third-year students after their 12-week clerkship in the department. Generated by the desire for a more reliable and valid assessment of clinical proficiency, the initial OSCE format was modeled after the work of Harden et al,¹⁸ with the clerkship committee identifying problems appropriate for measuring clinical skills. The examination format consisted of a circuit of 34 paired stations. At the first of each pair, students are required to perform a brief 4-minute interview or focused examination, and at the second of each pair, they record the data and answer questions about their patients and findings. A 90-second interval between encounters allows SPs to record observations about students (interview skills, questions asked or omitted, and examination technique) by means of a checklist. Since the time spent at each station is brief, students are asked to perform a very limited clinical task, such as "obtain the patient's history of present illness for headache" or "perform a physical examination of the patient's neck." In contrast to more comprehensive evaluation of single patients in other

Table 2.—Representative Stations of a Third-Year Clerkship OSCE*

Interview Stations Problem	Examination Stations	
	Problem	Examination
Fever and cough	Palpitations	Heart
Chest pain	Abdominal pain	Abdomen
Headache	Joint pain	Hands
Diarrhea	Neck pain	Thyroid
Weight loss	Health maintenance	Breast
Abdominal pain	Fever, confusion	Mental status
Dysuria	Leg weakness	Neurological
Rhinorrhea	Cough, weight loss	Chest and lung

*Multiple versions are created for each case, eg, chest pain scripts are written for both unstable angina and gastroesophageal reflux disease. OSCE indicates objective structured clinical examination.

exercises, the OSCE tests discrete skills in isolation, but samples these skills more broadly.

More recently (1989), the examination was expanded to 52 stations to evaluate students on an even wider range of skills. This updated examination involves 5 hours of testing, a duration usually associated with a higher degree of reliability than more limited evaluations.¹² A reliability coefficient of 0.8 is generally obtained with this examination format. Table 2 shows a representative sample of cases and skills measured with recent OSCEs.

A new clerkship in family medicine provided the opportunity for the most recent introduction of SPs in the curriculum. The format selected for the SP-based assessment is a focused encounter. As a summative exercise at the end of the 4-week rotation, a student performs a 30-minute interview and physical examination on each of two SPs. For each encounter the student is given a brief case scenario and a copy of the patient's "chart" shortly before beginning the exercise. A faculty member of the department observes each encounter through a one-way mirror. Upon completing the interview and physical examination, the student may ask for diagnostic test information (eg, laboratory values or x-ray reports), which are provided by the examination supervisor. The student is then given 1 hour to seek additional information from the library and consultants and to prepare the case for presentation to the faculty examiner. The student must also write a brief summary of the patient's problem. Case scenarios are derived from problems common in a family medicine practice, often focusing on disease prevention and ambulatory care.

The Senior Assessment Exercise: Evaluating the Product

Standardized patients are used extensively at UTMB in an evaluation exercise required of all fourth-year medical students before graduation.

Adopted by the faculty in 1987, the senior competency assessment exercise is modeled partially on the "triple-jump" exercise developed at McMaster University¹⁴ and the University of New Mexico's "individual process assessment."¹⁵ One notable difference between the UTMB exercise and its forebears is that the clinical encounter between student and patient is observed directly by a team of two faculty members: a clinician and basic scientist. Basic scientists are enlisted to observe the "product" they helped educate and to critique the student's basic reasoning and judgment. Each senior student performs a complete history and physical examination on an SP, writes a preliminary problem list, and is given several hours for independent study before reconvening with the pair of faculty examiners. Establishing a correct diagnosis is not an objective of the exercise. Instead, discussion focuses primarily on the student's approach to the patient, diagnostic reasoning, and the student's ability to apply basic biomedical knowledge to resolving the patient's problem or problems. Student participation in the exercise is mandatory and performance is judged on a pass or fail basis. A student whose performance is judged unsatisfactory is required to undergo a second evaluation. Because most students are observed with a single patient exercise, an inadequate sample for reliable assessment, this evaluation is viewed primarily as a screen of local candidates for the medical degree and, perhaps more importantly, as an opportunity for the faculty to observe directly the comprehensive clinical skills of our senior students.

Students typically appreciate the direct observation of, and feedback about, their clinical performances, and comment that such faculty involvement is rare in their clinical rotations. Students also react favorably to the exercise's emphasis on problem solving and decision making, rather than reaching a specific diagnosis. Student concerns match those of faculty participants, recognizing the difficulty in making general conclusions about student ability based on a single patient encounter. Students uniformly request that this form of assessment occur more frequently and earlier in the curriculum.

Faculty support is illustrated by participation of approximately 70% of medical school faculty each year. Clinical science faculty frequently comment on the discrepancy between their previous perceptions of a student's ability and the student's performance on the exercise. Basic science faculty have expressed uncertainty over their role, prompting

examiners to carefully select cases for their clinical and basic science applicability. For example, discussion of a case featuring a patient with a bleeding diathesis could assess a student's knowledge of the coagulation cascade and its clinical relevance to the patient's presentation. Both clinicians and basic scientists have responded positively to this "snapshot" of the graduate and the opportunity to work with faculty colleagues outside their discipline.

The Internal Medicine Residency: SPs in Postgraduate Medicine

Encouraged by the positive student and faculty feedback about the OSCE experience, the Department of Internal Medicine expanded SP evaluations to its house staff training program in 1984. First-year house staff (18 each year) are now evaluated each spring with a distinct 12-problem, 24-station OSCE, administered separately from the third-year student examination. This dedicated examination allows two new factors to be incorporated into the exercise. First, problem construction can be tailored to include greater emphasis on diagnostic and data integration skills, in addition to the more basic data gathering stations of the student OSCE, which emphasize interview and examination technique. Second, it includes the opportunity for examinees to repeat each patient station in the presence of a faculty member, receive feedback, and be allowed to confirm directly the correct findings or diagnosis. Although the shorter examination time limits this OSCE's reliability as an evaluation instrument (typically a reliability coefficient of 0.6), its educational value to the house officer is enhanced by providing direct feedback on areas of strength and deficiency.

AREAS OF INVESTIGATION IN SP-BASED INSTRUCTION

As the SP program has developed and matured at UTMB, initial concerns about recruiting and training persons willing to participate as SPs have resolved and additional areas of research and development have been initiated. The early goals of the SP program were to investigate the impact of SP encounters on student performance, and to explore how best to incorporate SP-based testing and evaluation in each course. Current interests include investigation of new uses for SPs in teaching and evaluation, motivational factors for SPs,¹⁶ and psychometric issues surrounding use of this methodology. To serve developmental, educational, and research functions, databases are compiled from checklists that SPs complete during

their portrayals. Accessing these databases for a given SP is helpful in identifying training issues (eg, one patient may be consistently stringent in grading student interviewing technique). Accessing databases by student or groups of students within courses can identify educational issues for faculty to address within each program (eg, all students were weak on palpating for the spleen). The breadth of the SP program at UTMB also allows development of research protocols to investigate longitudinal clinical competency issues. Research applications in the Introduction to Patient Evaluation course include investigation of demographic variables affecting student and SP performances (age of the SP, gender of patient or student),¹⁷ and issues of stringency biases in ratings. In Introduction to Clinical Medicine, to further clarify standards of performance, the course committee has investigated manipulations of scoring for SP encounters, and has prospectively identified 28 (29 for male patients) items considered critical to physical examination; omission of any six items results in a failing grade.¹⁸ This experiment to establish standards of mastery, a departure from typical normative-based assessments of clinical competency, is now in progress. Although precise performance standards based on faculty expectations may vary by institution, identification of specific criteria for competency at each school would allow students to be compared with an identified standard of performance rather than with each other.

The comprehensive format of the internal medicine OSCE (52 stations, 5 hours of testing) raised concerns about the effects of prolonged testing. Although student fatigue can be a factor with such long exercises, Petrusa,¹⁹ Lloyd et al,²⁰ and Rutala et al²¹ could not demonstrate that student performance was influenced by examinations of this length. In a cooperative study between Southern Illinois University, the University of Washington, University of North Carolina at Chapel Hill, and UTMB, the feasibility of building the same OSCE at four schools was demonstrated.²² Investigation of the OSCE as a predictor of residency performance is also under way, and preliminary results indicate a positive correlation of UTMB student OSCE performance with rankings of these students by our intern selection committee.²³ Early use of the same 34-station OSCE for students and residents allowed additional validation studies to be conducted. These analyses showed that performance by house staff in most areas of evaluation was superior to that of stu-

dents, and performance by second-year residents on higher order tasks (diagnosis and intervention), as well as total score was superior to that of first-year residents.²⁴ In addition, SPs were used in an American Heart Association grant to evaluate identification of cardiovascular risk factors in the internal medicine residents' clinic.²⁵ This application of SPs in an actual clinical setting allowed the investigators to assess residents' competence in recognizing and managing patients with cardiovascular risk factors without detecting the SPs. The preliminary results have generated interest in expanding this approach to assessing family medicine residents and faculty.

Analysis of the focused encounters during the family medicine clerkship has revealed no significant difference in student performance between cases, or based on time of year (ie, based on extent of clinical experience). Interrater variation has been found to affect scoring on the examination, however. This supports the findings of other investigators that evaluators often have differing expectations of student performance.¹² Data obtained from the senior assessment exercise have been used primarily as feedback to faculty on students' abilities, with student performance as a group reported to the faculty each year. Preliminary investigation into the diagnostic reasoning of students randomly assigned either to simple or difficult cases has also been undertaken and reported.²⁶

ISSUES IN SP-BASED INSTRUCTION

The use of SP encounters in clinical teaching and assessment at UTMB illustrates the adaptability of the technique to a wide range of educational objectives. Depending on the clinical tasks being measured, encounters may be tailored to focus solely on a chief complaint or organ-based physical examination.⁸ Alternatively, longer encounters allow a student to perform a complete history and physical examination for a global assessment of clinical skill.²⁷ The clinical tasks may also focus on areas underrepresented in general medical education, such as care of ambulatory patients, cost of medical care, and ethical issues.

A recent survey by Stillman et al²⁸ showed the extent to which SPs are used to develop and assess clinical skills in US and Canadian medical schools. Of the 136 schools responding, 70% now incorporate SPs in their undergraduate and graduate training programs. A few institutions have long had SP programs, but the average duration of 4 years indicates that most schools have only re-

cently adopted the technique. Some institutions use SPs early in a medical student's career, emphasizing development of history-taking and physical examination skills during the first and second years.²⁹ Others concentrate in the clinical years, with greater emphasis on integration of skills on a formative or summative basis.^{8,9,30} Still others extend use of SPs to their graduate medical training programs.^{7,24}

Standardized patient encounters are designed to simulate actual trainee-patient interactions with a high degree of realism. This goal is readily achieved when asking an SP to learn the script of a patient with chest pain and respond to student questions accurately and consistently. Recruiting patients with genuine and stable physical abnormalities adds to the realism of the exercise. However, simulation of some physical findings, such as an acute abdomen, can only be approximated in this setting. Each application must balance the advantages of in-depth evaluations of single patients with the greater reliability associated with measuring skills across many patients in a short station examination.

Issues of cost can affect a school's decision to incorporate SP-based techniques in its curriculum. Most SP programs are initiated, at least in part, to provide reinforcement and evaluation of clinical skills with less faculty effort, or to provide information about trainees that is not being obtained by other methods. Most institutions have difficulty in assessing clinical performance by direct observation because of conflicting priorities for faculty time. Standardized patient-based performance assessment can be a cost-effective addition to direct faculty observations. For example, a 5-hour OSCE at our institution regularly assesses clinical performance on 50 students in 26 skill areas, at a direct patient and administrative cost of \$1500. If presumed to replace the equivalent of 2500 faculty hours of observation time, this format can be a useful adjunct to traditional one-on-one teaching sessions.

The experience at UTMB and at other institutions suggests that all SP programs for teaching or assessment of clinical competency must address four major questions:

1. **Which clinical skills will be assessed?** A major strength of SP assessment rests in the faculty's ability to develop exercises that encompass a broad range of clinical behaviors, attributes, technical abilities, and decision-making skills (Table 3).

2. **What degree of skill integration is desired?** A clinical encounter may be

Table 3.—Clinical Skills Assessment With Standardized Patients

Category	Skill
Interviewing style	Professionalism, organization, empathy
Data collection	Question selection, patient evaluation technique
Recognition of pathology	Identification of critical patient responses, Discrimination of normal from abnormal findings
Differential diagnosis	Discrimination between different pathologic diagnoses
Intervention	Diagnostic test selection Therapeutic intervention Patient education Procedural skills

structured to allow students to perform a complete history and physical examination on an SP. This format is used more commonly as an educational rather than an evaluation tool because the longer encounter time limits the number of cases used and because of the high intercase variability associated with narrow item sampling.^{7,12} The multistation OSCE is a more typical application of SPs in clinical skills assessment, in which the individual components of clinical competency are assessed through focused history taking or physical examinations during brief (5- to 15-minute) patient encounters.³¹ Although each assessment station measures a discrete clinical skill in isolation, the brief encounters allow wider sampling of skills, providing greater reliability.¹²

3. What level of trainee will be assessed? Ideally, development of clinical competence should begin in the earliest days of medical training. For optimal feedback on strengths and deficiencies, competency assessment should begin early as well. Because clinical skills are developmental and typically learned in a stepwise process, SP-based education and evaluation can serve multiple roles—to provide a developmentally appropriate challenge, to reinforce positive behaviors, and to identify and correct deficient ones. Early incorporation of these techniques can help provide formative evaluation by identifying problems and providing immediate remediation. Most medical schools incorporate at least limited patient contact during early periods of training, providing an opportunity to teach, reinforce, and assess clinical skills at all stages of professional development.

Standardized patient encounters are very positive experiences for both preclinical and clinical students. The value

of these encounters may be enhanced by tailoring the sessions to the level of each examinee, emphasizing history taking and physical examination technique (first- and second-year medical students), pathology detection and differential diagnosis (third- and fourth-year students), and data integration and problem solving (interns and residents). The technique has also gained special attention from investigators of the NBME³² and the Education Council for Foreign Medical Graduates³³ as a possible adjunctive clinical competency assessment in medical licensing and certification decisions.

4. Will the encounters primarily serve teaching or evaluation roles? Independent of the skills emphasized, the format of the patient encounters, and the level of examinee experience, SPs may be incorporated effectively for either teaching or assessment. These encounters have been used successfully to identify deficiencies and reinforce clinical strengths before a course evaluation, particularly when combined with direct feedback to the trainee from patient or observer.³⁴ In its more typical application, the OSCE is used as an evaluation exercise after a clinical clerkship,⁸ before medical school graduation,¹¹ or as a competency assessment during postgraduate training.²⁴

If SPs are to be a superior tool to teach and assess clinical competency, one must be able to draw inferences about observed behaviors, and extend these conclusions to an examinee's behavior in the clinical setting. Reliability and validity are the issues affecting SP cases most directly. For a full description of these issues, the reader is directed to a recent comprehensive review by van der Vleuten and Swanson.¹²

FUTURE DIRECTIONS FOR INVESTIGATION

Much remains to be learned about the use of SPs in teaching and assessing clinical competency. Issues of cost and practicality must be addressed, and more research is needed on reproducibility of performance if pass or fail and licensing decisions are to be influenced by SP examinations. Nevertheless, addressing many of the unresolved issues of clinical competency measurement described here will require novel assessment techniques. Our experience, along with that of others, suggests that SP encounters may be particularly well suited to such investigations.³⁵

The evaluation of clinical competency at most medical schools ranks students or residents on a normative scale, ie, one examinee's rating is compared with another on an arbitrary distribution

scale. Although vague descriptors such as "above average" and "below average," or "competent" and "incompetent" may be used, these categories are rarely anchored by clearly defined behaviors.¹² Little is currently known about the impact of using a mastery testing approach to clinical competency, but SP encounters offer a unique opportunity to establish an absolute standard of performance based on direct observation of clinical skills. Investigators at UTMB are experimenting with standard setting exercises at several levels of instruction, some of which have been reported.¹⁸

Closely related to standard setting is the increasing emphasis on measuring outcomes in assessing effectiveness of an educational program. As accreditation agencies such as the Liaison Committee for Medical Education require greater accountability for the educational outcomes of medical training and competency of graduates, the stimulus for more objective measurements of clinical skills is likely to grow.⁶ Assessment of clinical skills with SPs at regular intervals, as practiced at UTMB, is uniquely suited to longitudinal analysis of examinee competence, permitting institutions to track students' performance as they progress through the curriculum.

Finally, educators have observed that evaluation methods can influence student study patterns, presumably because trainees view assessment methods as indicators of what teachers believe is important.³⁶ Greater emphasis on assessment of clinical skills has been shown in one institution to shift student priorities from textbook-based study to patient-based learning activities.³⁷ Testing with SPs permits teachers to prospectively identify critical skills for their trainees and establish explicit performance criteria for clinical competency. Revision of the NBME's certification examinations to emphasize synthesis of clinically relevant material over data recall is a recent example of a change in evaluation methods that may influence educational priorities. We believe additional emphasis on objective SP-based clinical competency assessment can achieve analogous changes in medical training. Along with other innovators in medical education, UTMB is prepared for these and future challenges in training the next generation of physicians.

The authors acknowledge Daniel C. Allensworth, MD, C. William Daeschner, Jr, MD, Andrew F. Payer, PhD, Barbara L. Thompson, MD, Barbara G. Ferrell, PhD, Michael Callaway, MS, and Karon Collins, MA, for their roles in standardized patient assessments at UTMB, and for their

contributions to this manuscript. We also thank the many faculty members who have developed and reviewed standardized patient case materials. We acknowledge the valuable editorial assistance of Alice Cullu.

References

- Association of American Medical Colleges. *AAMC Curriculum Directory*. 18th ed. Washington, DC: Association of American Medical Colleges; 1989.
- Stillman PL, Swanson DB. Ensuring the clinical competence of medical school graduates through standardized patients. *Arch Intern Med*. 1987;147:1049-1052.
- Bradford WD, Schofield JR. Study of required clerkships in internal medicine in US and Canadian medical schools. *J Med Educ*. 1986;61:157-162.
- Dauphinee WD. Clinical education: the legacy of Osler revisited. *Acad Med*. 1990;65(suppl 9):S68-S73.
- Association of American Medical Colleges. *Physicians for the Twenty-first Century*. Washington, DC: Association of American Medical Colleges; 1984:14-19.
- Kassebaum DG. The measurement of outcomes in the assessment of educational program effectiveness. *Acad Med*. 1990;65:293-296.
- Stillman PL, Swanson DB, Smee S, et al. Assessing clinical skills of residents with standardized patients. *Ann Intern Med*. 1986;105:762-771.
- Petrusa ER, Blackwell TA, Rogers LP, Saydjari CM, Parcel S, Guckian JC. An objective measure of clinical performance. *Am J Med*. 1987;83:34-42.
- Barrows HS, Williams RG, Moy RH. A comprehensive performance-based assessment of fourth year students' clinical skills. *J Med Educ*. 1987;62:805-809.
- Keynan A, Friedman M, Benbassat J. Reliability of global rating scales in the assessment of clinical competence of medical students. *Med Educ*. 1987;21:477-481.
- Stillman PL, Regan MB, Swanson DB, et al. An assessment of the clinical skills of fourth year students at four New England medical schools. *Acad Med*. 1990;65:320-326.
- Van der Vleuten CPM, Swanson DB. Assessment of clinical skills with standardized patients: state of the art. *Teaching Learning Med*. 1990;2:58-76.
- Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. *BMJ*. 1975;1:447-451.
- Painvin C, Neufeld V, Norman G, Walker I, Whelan G. The triple jump exercise: a structured means of problem solving and self-directed learning. *Proc Annu Conf Res Med Educ*. 1979;18:73-83.
- Kaufman A, Mennin S, Waterman R, et al. The New Mexico experiment: educational innovation and institutional change. *Acad Med*. 1989;64:285-294.
- Levine MK. *Medical Education: Factors Affecting Performance of Simulated Patients*. Santa Barbara, Calif: University of Santa Barbara; 1990. Dissertation.
- Rogers LP, Collins K, Callaway MR. Approaches for evaluating standardized patient performance in clinical assessment encounters. Presented at 10th Annual Conference for Generalists in Medical Education; October 30, 1989; Washington, DC.
- Dorsey NK, Petrusa ER, Kluge RM. Establishing a standard for the final clinical exercise in the introduction to clinical medicine course. Proceedings of the Third International Conference on Teaching and Assessing Clinical Competence; Groningen, the Netherlands; May 20-24, 1989:213-218.
- Petrusa ER. The effect of number of cases on performance on a standardized multiple station clinical examination. *J Med Educ*. 1987;62:859-860.
- Lloyd JS, Williams RG, Simonton DK, Sherman D. Order effects in standardized patient examinations. *Acad Med*. 1990;65(suppl 9):S51-S52.
- Rutala PJ, Witzke DB, Leko EO, Fulginiti JV, Taylor PJ. Student fatigue as a variable affecting performance in an objective structured clinical examination. *Acad Med*. 1990;65(suppl 9):S53-S54.
- Petrusa ER. Improving clinical performance assessment: a multi-institutional trial of the OSCE (symposium). *Proc Annu Conf Res Med Educ*. 1987;26:269-274.
- Blackwell TA, Dorsey NK, Ainsworth MA. Comparison of intern performance predictors. Proceedings of the Fourth Ottawa Conference on Assessing Clinical Competence; Ottawa, Ontario; July 7-10, 1990.
- Petrusa ER, Blackwell TA, Ainsworth MA. Reliability and validity of an objective structured clinical examination for assessing the clinical performance of residents. *Arch Intern Med*. 1990;150:573-577.
- Rettie CS, Ainsworth MA, Segal RC, Rogers LP, Sajid AW. The use of standardized patients in the assessment of residents' performance of preventive cardiology interventions. *Proc Annu Conf Res Med Educ*. 1989;28:153-158.
- Perkowski-Rogers L, Markus JF. Assessing clinical competency: case-specific analysis of a two-year project. Proceedings of the Third International Conference on Teaching and Assessing Clinical Competence; Groningen, the Netherlands; May 20-24, 1989:364-368.
- Williams RG, Barrows HS, Vu NV, et al. Direct standardized assessment of clinical competence. *Med Educ*. 1987;21:482-489.
- Stillman PL, Regan MB, Philbin M, Haley HL. Results of a survey on the use of standardized patients to teach and evaluate clinical skills. *Acad Med*. 1990;65:288-292.
- Vu NV, Steward DE, Marcy M. An assessment of the consistency and accuracy of standardized patients' simulations. *J Med Educ*. 1987;62:1000-1002.
- Stillman PL, Regan MB, Swanson DB. A diagnostic fourth year performance assessment. *Arch Intern Med*. 1987;147:1981-1985.
- Harden RM. What is an OSCE? *Med Teach*. 1988;10:19-22.
- Swanson DB, Norcini JJ. Factors influencing the reproducibility of tests using standardized patients. *Teaching Learning Med*. 1989;3:158-166.
- Conn HL, Cody RP. Results of the second clinical skills assessment examination of the ECFMG. *Acad Med*. 1989;64:448-453.
- Hodder RV, Rivington RN, Calcutt LE, Hart IR. The effectiveness of immediate feedback during the objective structured clinical examination. *Med Educ*. 1989;23:184-188.
- Miller GE. The assessment of clinical skills, competence, and performance. *Acad Med*. 1990;65(suppl 9):S63-S67.
- Newble D, Jaeger K. The effect of assessments and examinations on the learning of medical students. *Med Educ*. 1983;17:165-171.
- Newble DI. Eight years' experience with a structured clinical examination. *Med Educ*. 1988;22:200-204.