STATUS OF STANDARDIZED PATIENT ASSESSMENT

Taking Standardized Patient-Based Examinations to the Next Level

Emil R. Petrusa
Associate Dean for Curriculum Assessment
Duke University School of Medicine
Durham, North Carolina, USA

Purpose: The purpose of this article is to summarize this author's view on "where we are" with standardized patient-based assessments of clinical performance and to offer three directions for further research and development.
Summary: *The push for more objective outcome data has fueled proliferation of the most researched innovation in the history of medical education. Near-random clinical experiences of students do not provide consistent, repeated practice with important clinical cases to achieve minimally adequate performance on these objective performance examinations, leading to scoring "psychogymnastics" to titrate fail rates.*
The second area is to modify these examinations to reflect features at higher levels of professional development such as situational awareness. Theories of professional development should guide changes.
The third area incorporates multiperson scenarios; a clinician with a family or a team in the operating room. Simulation of complex situations, especially those requiring rapid, accurate communication and action can reduce medical errors and improve patient safety.
Conclusions: *Standardized patient-based examinations provide objective outcome data but require artificial adjustments in scoring due to inconsistent learning opportunities. Theoretical research on professional development, acquisition of expertise and team functioning provides fertile, new directions to take standardized patient-based examinations to the next level.*

Teaching and Learning in Medicine, *16*(1), 98–110

The purpose of this article is to succinctly summarize the view of "where we are" with standardized patient (SP)-based assessment of clinical performance and to offer directions for further research and development. The reader is directed to several detailed reviews[1–3] of this assessment approach that have appeared in the literature over the last decade. Only general reference to the conclusions and recommendations in these excellent reviews are made here. After a brief overview of the current state of affairs, this article elaborates three major areas where exciting new development is needed. The first area concerns the gap between SP-based assessments and the current clinical curriculum of medical schools. The second focuses on advances in these examinations that correspond to performance at advanced stages of professional development. Finally, the third area builds on features of the highest levels of professional development but goes beyond the traditional clinical dyad of doctor and patient; beyond the tasks of diagnosis and management, to more complex situations such as a single provider who must manage several patients simultaneously or where a leader must coordinate the work of others in a changing situation. Literature in aviation and military training offer constructs and measurement tools, for example, the construct of "situational awareness" that will be immediately useful in assessing individual and team performance in complex clinical situations.

Correspondence may be sent to Emil R. Petrusa, PhD, Associate Dean for Medical Education, Duke University Medical School, P.O. Box 3628 DUMC, 600 Davison Building (Green Zone), Durham, NC 27710, USA. E-mail: petru001@mc.duke.edu

Current State of Affairs With SP-based Examinations

The proliferation of SP-based examinations of the last 20 years is evidence that these tools are perceived as useful in medical education. Research on measurement characteristics of these tools has been summarized previously.[1–3] Evidence supports the accuracy of SPs, both their portrayal and marking of performance in the exam. There is no evidence for a gender bias by either the SPs or examinees.[4,5] There is some evidence to suggest that encounter length influences how examinees choose to use their time with the patient.[6,7] Very short encounter lengths correspond to a high control, targeted series of questions presented to the patient. Longer encounter times allow a more patient-centered, less directed interaction. Cases are built around specific behaviors expected from students. These expected behaviors become the checklist completed by the SP. Case formats almost invariably require a focused workup. Often there are post-encounter questions the examinee must answer. Post-encounter questions may be of multiple choice, true–false, matching or short, free-response format. Guidelines for assigning points to the free-response answers increase the consistency of scoring. Whereas the post-encounter tasks seem directly relevant to the patient just seen, earlier work suggests answers to post-encounter questions lower overall reliability of the assessment.[8,9] Therefore, the SP-based exam is objective, standardized, and consistent in assessing the performance of all students. Predetermined pass marks for cases or the whole test are applied equally to all performances.

Pass marks are set in a variety of ways. Some are based on points (or percentage of points) accumulated across a series of patients. The pass mark may be a fixed percentage (e.g., 70%) or a position relative to other students (e.g., –2 *SD*s below the class mean). Other approaches set a case-level pass mark, and then the full examination's pass mark is based on the number of cases passed. Others use a mixed approach where passing a certain number of cases is one mark and performing adequately on the doctor–patient relationship is aggregated over all cases.

When initially reported, standardized clinical performance examinations (CPXs) included one or two patients in a circuit of 10 or more stations.[10–15] Subsequent adaptations had a complete circuit comprised only of patient interactions, some simulating diseases, others with stable clinical findings.[16,17] Further adaptations included longer encounter time and varying the kinds of post-encounter activities.[18,19] Based on substantial research, SP-based examinations now are part assessments for licensure and certification in Canada.[20] Persons trained outside the United States who wish either to practice in the United States or to pursue additional residency training must pass a 10-case examination in addition to a written test of clinical knowledge.[21,22] The United States Medical Licensure Examination is scheduled to include an SP-based examination as a complement to the Step 2 test of clinical knowledge.[23,24] The accreditation process for North American medical schools requires more objective data to support the claim that by graduation medical students have achieved the goals and objectives of their medical school.[25,26] The push for more objective outcome data has fueled proliferation of SP-based examinations.

SP-based examinations are arguably the most extensively researched educational innovation in the history of medical education. In a recent effort to catalogue all published work on SP-based examinations, one research group has identified over 800 articles.[27] With such a pervasive literature base it is reasonable to question whether this assessment approach has reached its evolutionary pinnacle. Are there any more important research questions to be studied? This article not only intends to answer with a resounding "yes," but also suggests directions for key studies, some that will improve conceptual and psychometric aspects of existing exams and others that will lead to exciting new applications. The vast majority of SP-based assessments occur throughout the curriculum of medical schools. This pervasive use warrants a closer analysis of its role in a medical school's educational system.

Saving Face (Validity)—The Role of SP-Based Examinations in Medical School Curricula

As mentioned, accreditation of North American medical schools calls for more objective outcome data. Checklists, predetermined processes for aggregating data, and guidelines for scoring written responses rightfully qualify these exams for the objectivity desired by the accreditation body. I contend that medical schools have supported these exams, in part, because they can save face during the accreditation site visits. However, careful analysis of the internal structure and their role as one of the three key components of an educational system will reveal limitations that undermine the validity of the claim of "objective outcome measures." To set the context for this analysis, it will be instructive to briefly review key features of current clinical curriculum of medical schools where a curriculum has been defined as " … structured, planful, learning activities … "[28]

A sound educational system has three components: (a) goals and objectives that indicate what is to be learned and achieved, (b) activities organized to facilitate learners accomplishing those goals and objectives, and (c) an assessment of learners to see if they have

achieved the goals and objectives as planned. If a sizable number of learners do not meet expectations, analysis of the three components attempts to uncover the reasons for this lack of achievement. Problems with one or any combination of these three components could result in low performance.

Most clinical curricula develop students' clinical skills by first teaching them to take a comprehensive history and conduct a complete physical examination. In medical interviews or an introduction to clinical medicine-type courses students are taught all of the elements and are evaluated on their ability to reproduce all of the elements. Therefore, to do well a student needs to ask all of the questions and execute all of the examination maneuvers. Even during clinical clerkships (usually the 3rd and 4th years in North American medical schools) students are expected to conduct "complete histories and physicals" on patients seen in the hospital. In ambulatory settings, students may begin to focus their workups on the patient's chief complaint or reason for the visit. This is due, in part, to the shorter timeframe for seeing patients. However, even in ambulatory situations, particularly subspecialty clinics, students still may be expected to do comprehensive workups. Neither in hospitals nor clinics is there any prescribed curriculum of patients that all students must competently workup.

Educational goals for clinical clerkships include something similar to "the student will be able to take a history and conduct a physical examination on (for example) pediatric patients." What constitutes an adequate history and physical examination? Certainly one can imagine an inadequate history or physical examination. What the clinical curriculum needs to provide is a description of what is "good enough." This description will change with different clinical information. A good enough history of a child with chronic diarrhea will be different from one for a child with a persistent cough. The range of possible etiologies that should be explored for each problem is different; therefore, the historical information, physical examination, and diagnostic studies for each will be different. Consequently, the actions that constitute good enough for the two situations will be different. Certainly, experienced clinicians are capable of judging good enough for these two situations. However, this is often implicit, not directly revealed to the student. Further, two attending physicians may have considerably different notions of what is good enough for a given patient. Finally, no student sees the same patient as other students, and each student is guided in their development by more experienced practitioners who have different ideas of acceptable performance. The educational challenge is to assure that all students receive a true curriculum of clinical problems. The assessment challenge is to provide exercises sampled from the curriculum that allow consistent judgments about the scope, depth, implicit logic, and accuracy of each and every student's clinical performance at least with the clinical challenges explicitly taught in the clinical curriculum. Without a consistent clinical curriculum, it is not practical to know what patients each student has experienced and what level of performance each attending physician expected. Without this information, only the lowest levels of performance may be expected on standardized, objective assessments of clinical knowledge, reasoning, and performance. Even with only the lowest levels of achievement expected, numerical "adjustments" of scoring and pass marks on any standardized clinical examination will continue to be necessary so that the appearance of adequate achievement by most students is maintained.

Descriptions of patients in the earliest objective structured clinical examinations had a real or simulated patient and one or two faculty observers that evaluated an examinee's performance.[10,13] Observers used checklists to document the behavior of each examinee, hence the word *objective* in the name, objective structured clinical examination. In the mid 1980s with the advent of an "all patient" standardized examination,[16,17] developers continued use of the behavioral checklist but put them in the hands of the SPs. A 17-patient examination given three times per year in an internal medicine clerkship with 68 students per rotation would have required the department faculty to observe and evaluate over 1,156 SP-examinee interactions. It was decided that this was not feasible. This left the task of recording an examinee's behavior to the SPs. The only conceivable approach was for the SP to record whether or not a student performed any of a limited set of expected behaviors. It was reasoned further that student answers to questions after seeing each patient would indicate their understanding of a patient's problem. The hypothesis (still not fully tested) was that these post-encounter questions would validate the student's thinking reflected in their actions during the encounter. Checklists completed by SPs and post-encounter questions to assess an examinee's understanding remain the most common approach to collecting information about an examinee's performance to the present time. Face validity was high—these encounters approximated components of real patient workups—and the post-encounter questions looked like those asked of students during case presentations in the hospital. Performance with each case is evaluated the same way for all students—considerably more objectively and consistently than ratings from residents and attending physicians who work with a student for 1 month, but rarely observe an actual workup.

Although scores, usually expressed as a percentage of expected behaviors, vary with a particular case, the average score on these examinations is between 65 and 75% (with a standard deviation of about 10% of this,

i.e., 6.5–7.5). This level of performance is so consistently reported as to be almost a universal truth. In traditional grading schemes 70% is usually the borderline between passing and failing, yet this range is quite acceptable for standardized clinical performance assessments. Why is this a problem?

Competent clinical performance is grounded in correct actions. Malpractice suits often hinge on a single behavior, whether it is a wrong action taken or a correct one omitted. The scoring scheme of a standardized clinical performance assessment is also grounded in correct behaviors. However, students are only getting credit for about 70% of what is expected. Whereas only doing 70% of what is correct may constitute malpractice for a licensed physician, this level of performance may be quite acceptable for students. It is my opinion that this acceptance reflects the mismatch between the nearly random clinical experiences of medical students and the particular cases in these assessments. Evidence to date is that the typical clinical experiences are not sufficient to educate students to perform competently even for relatively straightforward problems such chest pain, abdominal pain, and low back pain. Even if medical schools engineered patient encounters so that each student had a one-time exposure to these problems, students would still perform well below a clinically acceptable level. Research on the acquisition of expertise indicates that many, many occurrences are needed, coupled with corrective feedback, to approach competent, let alone expert, performance. Although SP-based examinations are more objective than other assessments of clinical performance, considerable adjustments in scoring must be made to compensate for current clinical curricula. Whatever level of performance is demonstrated by a class of students, that performance must be declared acceptable unless and until there is a change in the methods of clinical curriculum. Such allowances and other scoring manipulations would not be necessary (read: it is hypothesized that) if clinical education were restructured. Although space limitations prohibit a detailed description and rationale, some general educational features should provide a sense of the type of changes recommended.

Research to address this issue of 70% performance on standardized CPXs should take two paths. One path is to investigate the typical acceptable level of performance for a minimally competent encounter. It is likely that the clinical faculty members who created the checklists and post-encounter questions are overly optimistic and idealistic about the behaviors necessary to competently respond to a patient's problems. There is strong evidence that faculty overestimate the performance of examinees when setting pass marks for written examinations.[29] The second path for research is the type, amount, and frequency of specific practice with particular problems that will result in improved performance on these examinations. The first path assumes current clinical education is adequate and the problem lies with the scoring of the examination. The second one assumes that the performance expected on the examinations is desirable and the problem lies with the structure of learning clinical medicine. Almost certainly the optimal outcome will derive from changes in both. The Association of American Medical Colleges has a strategic initiative to improve clinical education, especially for medical students.[30] Several recent documents highlight selected medical schools' innovative changes to their clinical education. The need to alter the clinical experiences of medical students is clearly realized.[31–33] Whatever new form these experiences take, repeated practice on specific problems and clinical challenges, coupled with accurate and timely feedback should be a key feature of any clinical curriculum.[34]

As mentioned earlier, SP-based examinations fit the expectations of the accreditation agencies for objective data as well as any other approach for assessing clinical performance. In part, this explains their widespread use. When viewed as a part of an educational system, however, the "fit" is not so clear. It is not so clear first because all reported levels of performance in these assessments do not reach clinically adequate levels, yet no mention of this shortcoming is addressed. The second indication of a suboptimal fit is that there has been no report of systematic increases in level of performance on any SP-based examination, some having been administered for nearly 20 years.[35–37]

Let us analyze the relation among the three educational components. Because the clinical curriculum has no specific patients that must be seen, indeed, students often say that they have never seen one or more of the SP cases; how should students have learned how to focus workups? Expected performance with real patients is a comprehensive history and physical examination, yet the SP-based assessments are structured for focused workups. Further, a competent focused workup with a given patient, at least at the level of 3rd and 4th year medical students must include the logic and organization of the data collection around increasingly refined differential diagnoses. However, SPs simply record whether certain actions, listed on the checklist, occurred. SPs are not trained to assess logic and organization of data collection for their cases, so these important aspects of a student's performance are not assessed—a serious omission in the assessment of relatively advanced clinical performance. Without repeated practice and feedback on a set of clinical cases, indeed without any content-specific curriculum, the level of expected performance must be "adjusted" so that some, but not too many students fail the performance examination. Finally, if these objective, structured clinical examinations were truly regarded as one of the three fundamental educational components, then test-wise scores should increase over several years, as a school's faculty adjusts the curriculum in response to

moderate levels of performance. However, the 20 plus-year literature is curiously consistent with average percentage scores hovering around the 70% mark regardless of specific cases, a school's research reputation, or a problem-based learning curriculum.

A single study[38] is a candle in the evidentiary darkness of effective clinical curriculum outcomes. Students who had specific instruction on doing a mental status exam and three other neurology problems at one of four teaching hospitals did significantly better on the cases of dementia and gait disturbance in the end-of-clerkship structured clinical examination than students who rotated through other teaching hospitals that had no structured curriculum. Specificity of this impact was further demonstrated when students who experienced the structured curriculum for these two clinical problems did not do better (or worse) than their colleagues on other cases in the exam. For clinical performance to improve, clinical education must be much more focused to a particular curriculum, and learners must receive feedback with opportunity to repeat that performance until it reaches the desired level.[39,40]

What are the consequences of a near-random clinical curriculum and a case-specific performance examination? First, students justifiably will feel trapped by the school in that quality performance is expected on a high-stakes examination, but little or no instruction is provided so that students have an opportunity to acquire the skill necessary to do well on that examination. Second, scoring to determine what constitutes doing well will go through "psychogymnastics" to assure that some, but not too many, students fail. Neither normative nor criterion referenced approaches is immune to quantitative contortions.

Performance on cases may be scored as "percentage of correct actions/response" and a pass mark determined for each case (e.g., 70% or correct or –2 *SD*s below the class mean). Pass marks might be set only for aggregate performance across a set of cases. Again, passing the test might require achieving 70% of all actions and responses to post-encounter questions or achieving higher than 2 *SD*s below the class mean on total exam points. Although percentage of points reflects "more is better," this approach allows especially key actions or responses to count the same as less important ones. Eliciting rebound tenderness in a patient with acute abdominal pain counts the same as asking if anything makes the pain better. Test-wise percentage scores mask unacceptable performance on some, perhaps more important, cases. Getting 98% of the points for the depression case and 42% for the chest pain case averages to 70%. Percentage of case points or percentage of test points notwithstanding, it is the proportion of students failing the performance examination that matters most. If more than about 10% of a class fails this exam, the exam will be suspect. Too few failures and the exam becomes "too expensive." Titrating this all important fail rate may require adjusting the case pass mark to a particular score, such as 67% or 78% or to a relative position like –2.3 or only 1.67 *SD*s below the class average. None of these gymnastics lead to more competent graduates of our schools.

This issue remains even if scoring is based on specific behaviors (i.e., criterion referenced). The North Carolina Medical Schools Consortium has been administering a common set of cases to end-of-clerkship students since 1993. In 1994, adequate performance on a case was defined from the perspective of "minimal acceptable clinical practice." Clinical faculty members identified case-specific actions that they considered to be minimally clinically acceptable. The intent was to ground case pass marks in a "clinical standard of practice." Due to the lack of alignment between the CPX and the clinical curricula at the four medical schools, adjustments of the scoring were needed to keep the percentage of students that failed at a reasonable level. A similar outcome was found when this approach was used for setting pass marks for an examination given to residents.[41] These psychogymnastics are illustrated with the case of a man with a cough.

"Juan Carlos Hernandez" is seeking medical assistance for a persistent, dry cough. Mr. Hernandez had moved his family around the country as needed to obtain work but has lived in North Carolina for 5 years since getting a job on the paint crew of the medical center. He speaks broken English. Students have 15 min to work up Mr. Hernandez and then 10 min to answer questions about his case. The Appendix shows the history and physical examination actions expected of a student and questions to answer after the encounter. Bolded items are considered "critical actions" or the clinical standard of care for Mr. Hernandez. Table 1 shows the percentage of students in the consortium that received credit for each critical action in a recent administration (summer and fall of 2002). Despite nearly 95% of students listing tuberculosis in the top three of the differential diagnoses (data not shown), two thirds did not ask the patient to cover his mouth while coughing, and nearly one half did not list lung cancer in the differential diagnosis. Table 2 shows the impact of "forgiving" from zero to five critical actions on the percentage of students passing the Hernandez case. If held to a standard of minimal clinical competence (0 critical actions forgiven), only 14% of students completing 12 months of clinical experiences would have performed this well.

The examination pass mark was set through a method that incorporates both normative and criterion-referenced standards.[42] Faculty judgments indicated that no more than 10% to 12% of students should fail and that they should fail no more than 4 of 15 cases and still pass the test. Each case counts equally toward this total. The Appendix also shows the percentage of students who would have failed the examination under

Table 1. *Percentage of Students Not Doing Each Critical Action for Hernandez Case*

Critical Action	Percentage of Students Not Doing (%)
Establish good doctor–patient relationship	1.6
Communicate well	0.5
Asked about duration of cough	0.5
Asked about waking at night with sweating	15.0
Asked about loss of appetite and/or weight loss	14.1
Requested patient cover mouth when coughing	67.6
Listened to chest with stethoscope	3.8
Listened to back with stethoscope	8.7
Listed TB in top three DDX (of six)	6.0
Listed lung cancer in DDX	44.9
Ordered chest x-ray	1.6

Note: $N = 185$. TB = tuberculosis.

the six scoring conditions for the Hernandez case alone. Scoring for all other cases was unchanged. In the far right-hand column of Table 2, one can see that the percentage of students passing the whole CPX varies with the changes in the Hernandez case alone.

If the pass mark for each case were the clinical standard of practice (all critical actions performed) and students had to pass at least 11 of 15 cases, no student completing their clerkship year would have passed this examination (see Table 2). Additional data in Table 2 shows the impact of more lenient case and test pass marks for this same group of students. Forty-five percent of the students would have passed the CPX if one critical action were forgiven for each case and they had to pass 10 of 15 cases. If the pass park of 11 of 15 cases were maintained, then only 30% would have passed. By allowing two missed critical actions per case, 81% would have passed at least 11 cases. If a medical school allowed three critical actions to be missed on every case, it could claim an "outstanding success" of 91% of their students passing 80% of the CPX cases. The North Carolina Medical Schools Consortium allows from one to three critical actions to be missed for the 15 cases. This typically results in a pass rate between 88% and 92%. Clearly, these medical schools have not aligned the clinical curriculum with the cases in the CPX. The "take home" message from Table 2 is that there are multiple ways of setting the pass mark for SP-based performance examinations that allow the claim of success as long as the details about the pass marks are not fully disclosed. Unfortunately, some kind of psychogymnastics will be needed until the proper changes are made to the clinical curriculum. If medical schools were to require performance at the minimal clinical standard of practice (one would like to think schools would want to uphold this level), then much work is needed during the clinical years of medical school to assure that graduates can function at this level, at least for these common and fairly straightforward cases, as they enter post-graduate training.

Table 2. *Percent of Students Passing Hernandez Case for Various Allowances and Impact on CPX Pass Rate*

Allowance for Critical Actions Not Done[a]	Percentage of Students Passing Case[b] (%)	Percentage of Students Passing CPX[b] (%)
5	100	89
4	99	88
3	95	88
2	80	83
1	49	79
0	14	73

Note: CPX = clinical performance examination.
[a]$n = 11$. [b]$n = 185$.

SP-Based Examinations in the Context of Professional Development

This second area for advancements in SP-based examinations is closely related to the first. Certainly, alignment of assessment and curriculum is a key feature of a quality educational program. Over the course of formal medical training, learners should become more efficient and more accurate in their workups. Research on clinical reasoning in the early 1970s clearly indicated that expert clinicians working up problems in their area of expertise have the eventually correct diagnosis in mind earlier and require few pieces of information to confirm their diagnosis than competent clinical faculty members not expert in the area of the patient's problem.[43] If these clinicians were assessed in an SP-based exam where percentage of points was the scoring approach, the experts would likely have lower scores than the nonexperts. To claim validity for assessing clinical performance, scoring needs to be based on critical actions perhaps with a "bonus" for more efficient workups that included all of the critical actions. Simply doing fewer actions is not the same as doing only the most critical.

At pre-expert levels of training, we may not expect the most efficient and accurate workups, but we will want to know if the learner is "approaching the problem" adequately. A simple percentage of points will

Table 3. *Proportion of Students Passing CPX at Various Case aPss Marks*

	Number of Cases of 15 That Must Be Passed to Pass the CPX															
Number of Critical Actions That May Be Missed Per Case	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
0	100%	99%	92%	82%	61%	37%	21%	12%	8%	4%	1%					
1						100%	96%	90%	79%	64%	45%	30%	15%	7%	1%	0%
2								100%	99%	98%	93%	81%	66%	41%	15%	0%
3											100%	98%	91%	75%	42%	0%
4												100%	99%	92%	57%	0%

Note: $N = 194$. CPX = clinical performance examination

not tell us whether the workup was logical and organized around an increasingly specific differential diagnosis, yet this is the method used in most all SP-based assessments. Empirically, we may find that "doing more" with a patient does correlate with better performance, even without a direct measure of logic and organization. Several groups are experimenting with ratings of more global features of performance.[44–46]

Concomitantly with increasingly more efficient and accurate data collection, developing clinicians should conceptualize problems more similarly to expert conceptualization. Bordage and colleagues[47–49] identified five stages of understanding clinical data. At the lowest level of conceptualization, learners simply attempt to hold individual pieces of clinical information in their heads. When asked to describe the patient's problems, the learner reports each piece of data in a dispersed, disorganized manner. At best, the organization of this presentation follows some nominal structure such as "head to toe," "history of present illness followed by complete review of systems," or subjective and objective. At the highest level of understanding, the learner integrates data into dimensions or characteristics that discriminate categories of problems, such as "a slowly progressive onset" or "regularly irregular rate and rhythm." Empirical work is needed to develop a consistent measure of conceptual understanding through post-encounter tasks. Once a reliable measure is established, studies should look at the relation between the amounts and kinds of data collected by learners who have different levels of understanding about a given problem. It is conceivable that this line of investigation would render history and physical exam behaviors minimally important, at least for assessments at the level of clerkships and beyond. Later in training, focus will shift to interpreting the answers, but for now simply demonstration that the student knows all questions is sufficient. Learning the physical examination is analogous.

Throughout the curriculum, students demonstrate their competence by asking a complete set of history questions and conducting a head-to-toe physical examination on actual patients. The implicit educational philosophy seems to be that students must be thorough early in their professional training so as not to omit critical data. Later, with more clinical experience, orders will have associated certain patient information with particular problems or specific diagnoses. At this point, history taking and physical examination should become focused, that is, related to the patient's particular complaint or reason for hospitalization. Experienced clinicians are both efficient and more accurate in the data they seek and in its interpretation.

At this expected level of performance, assessments must attend to these features of logic and organization. The use of clinicians either as the SPs or as observers is the most efficient approach. It is not outside technical capability, however, for "smart systems" to be developed that would apply protocol patterns to the sequence of examine the behaviors. In other words, computers now have the capability to both interpret the spoken word and to determine if it in the pattern of portions to that of expert clinicians. At the present time, this seems too expensive. Experienced mid level practitioners should be quite able to serve as both SPs and assessors. Another opportunity is to use upper level residents in this capacity. A combination of portrayal and assessment would fit nicely into a seminar, workshop, or longitudinal experience intended to develop residents' educational skills. Because a small portion of residents eventually become junior faculty members, the investment of their time in this way will be returned in the form of more educationally advanced faculty members.

Possible mechanisms for assessing conceptual understanding include having the learner present the case to a clinician (resident, fellow, faculty member, physician assistant, nurse practitioner) who would score the responses. Another approach is to have students dictate an assessment of the patient's problem. Nonclinicians might use scoring templates with prototypical examples for various levels of understanding. Still, further work might identify similar dimension-like descriptions of management options, such as "controlling blood pressure while simultaneously decreasing blood thinners and maintaining potassium levels" similar to

those representing levels of diagnosis understanding. Such work could lead to modification of SP-based examinations for residents and practicing clinicians where the clinical challenges are on the management more than the diagnosis side of caring for a patient. It is quite possible that this line of research would lead to a conceptualization of developmental stages of understanding management, similar to Bordage's for understanding diagnoses.

Finally, theories of professional development, such as those by Dreyfus and Dreyfus[50] and by Eraut,[51,52] describe advanced levels of expertise as "grasping the whole situation," "intuitively incorporating situation-specific parameters into decisions about what to do." Situational awareness is a key predictor of expert performance of fighter pilots,[53] anesthesiologists[54] in emergent situations and multilevel military groups.[55] Supervisors and peers can provide reliable ratings of situational awareness; self-ratings are less reliable.[56] Clinical examples that would allow for situational awareness might include "determining alternative approaches to diagnosing Mr. Smith's problem because he has no insurance," or "let's get Mrs. Jones seen by anesthesia today because she has to rely on neighbors to get her to the doctor," or "we're short handed today so I (the resident) will draw this patient's blood, while you (the student) bring us something from the cafeteria—we're going to have to work through lunch." This last example reflects situational awareness in the context of teamwork. The third area for development of assessment in structured simulations is assessment of multiperson interactions in emergent and nonemergent situations.

Simulation-Based Assessment Beyond the Clinician–Patient Dyad

With a few notable exceptions, the vast majority of encounters in SP-based examinations are dyads consisting of an examinee and a SP. This is appropriate because most clinical interactions take that form and this is the first level of health care complexity for learners to accomplish. New cases should be developed to challenge more experienced clinicians. In addition to ambiguous signs and symptoms that may relate to several etiologies, these new cases should present challenging emotional states, socioeconomic and cultural differences, and management complications created by a patient's use of alternative medicine remedies. Certainly, these developments will occur as accreditation of graduate medical education presses for more objective data to indicate that residents have achieve goals and objectives at that level. The really innovative work will be the development of structured simulations that allow assessment of multiple persons in complex interactions (e.g., teamwork).

Analysis of near misses and other airline mishaps indicates that the majority of errors occur "within the system of interactions" rather than with any individual's technical competence.[57] High fidelity simulation of a patient in the operating room provides anesthesiology residents to face crisis situations in a safe environment.[58] Debriefing with a videotape of the resident's actions in the crisis allows the resident to see and correct less effective actions. A variation on the same crisis may be rerun so that the resident can practice better responses. In more complex situations, a whole surgical and anesthesiology team must deal with a crisis. Videotape replay along with checklist and rating data[59] allow analysis and correction of actions and cognitive processes by any or all participants. Inserting a person who will create difficulty in a predicable way during the crisis (e.g., a standardized circulating nurse or first assistant or an anesthesiologist) creates further complexity. A similar pattern is emerging from detailed studies of medical errors.[60]

There are already reports of multiperson simulations in the literature. The simulated family is a scenario where the medical student encounters parents, children, and a grandparent over 16 weeks during which different family members are seen for acute, chronic, and emergent problems and the death of the grandfather.[61] One school is conducting multidisciplinary scenarios where students from medicine, dentistry, public health, and allied health work on the coordinated care of patients.[62] These scenarios are used formatively, more for education than assessment. The conceptualization of proper roles and participation in the planning for and care of a patient provide the basis for development of appropriate assessment tools. These scenarios call for assessment tools that address leadership, cooperation, collaboration, and communication as well as the special knowledge from each participating discipline. In the present form, each student from each discipline participates spontaneously; there is no standardized role player, other than a "paper patient." An innovative use of scenario-based simulation puts preceptors in a circuit of ambulatory teaching situations. Each "station" in the circuit was a different teaching situation that might occur in an ambulatory setting. Standardized students portray learners with different challenges for the preceptor.[63]

The airline industry and anesthesiology use uncommon, but authentic, scenarios of possible disaster to give pilots and operating room participants experience with rapidly developing emergency situations that require precise, rapid communication, acute situational awareness, leadership, and collaborative decision making. These scenarios were developed for use with a life-like mannequin that simulates a patient under anesthesia. Scenarios range from an unanticipated physiological response by the patient to a fire in the operating room.[58] Operating rooms scenarios have included a

standardized surgeon, nurse, or anesthesiologist whose actions are designed to increase the tension and complexity of the situation. Typically, the interactions are videotaped and then replayed immediately after the scenario to assist participants in assessing their behavior. The same or similar scenario can rerun so that participants can "get it right." Ratings scales have been developed for use by observers to assess dimensions of interaction that relate to the group interaction as a whole.[59] More development work is needed for tools that capture the thoughts and feelings of the various participants at key points during the scenario.

Other health care situations where simulation with reflection and feedback would be useful are (a) running a code, (b) simulating an emergency department with multiple traumas, and (c) managing a team of health care professionals (e.g., a hospital ward service) where simultaneous and interacting goals of patient care and education should be accomplished. Running or participating in a code is a standard component of certified Advance Cardiac/Trauma Life Support courses. Those scenarios occur within the course where skills, situational awareness, and other participants are in the forefront of consciousness. The application of those skills, awareness, and management at an unexpected place and time is better for determining the effectiveness of training because cuing is reduced. A scenario might be orchestrated in the students' lecture hall or in the cadaver lab. Careful and detailed planning, similar to that for introducing SPs into a real clinic,[64,65] will assure the scenario is contained either by stopping the scenario prior to activating the actual alarm or informing the regular response system that a simulation is occurring. Digital video cameras are small enough to be unobtrusive and portable enough to capture the situation for future reflection.

A clinical skills laboratory could be the setting for an emergency department scenario where students (or with residents and faculty members) must rapidly process progressively complex situations with triage priorities that change during the scenario and where some confusion or misunderstanding alters expected events. The complexity and difficulty of the scenario can be relatively straightforward or chaotic, depending on the purposes for running the scenario. Again, videotape review with sensitive and specific assessment tools will increase the effectiveness of learning.

Team and multiple agency response to emergent events in the community is another useful application of simulation methods. "Usefulness" is assumed under the premise that specific practice, with feedback, will better prepare responders than the accumulation of experience with situations very different from these special circumstances. Scenarios of a hostage situation in a medical center building where SWAT personnel, emergency medicine technicians, and emergency room physicians must work together to subdue the perpetrators and respond to the injured has been run successfully at the University of Miami.[66] "Standardized citizens" portray normal building staff with various injuries and psychological trauma. This group has funding to develop and run scenarios in the community where weapons of mass destruction, including chemical and biological weapons, are involved. emergency response agencies have not practiced situations where a relatively large number of people are possibly exposed to chemical or biological agents. Rapid decisions about the proper protective equipment, about containment and isolation of contaminated citizens, and about management of exposed emergency personnel just part of the simulations. Coordination with hospitals, police, fire, rescue people, and the various community agencies, such as the civil defense, government, and media may be added to the scenario for realism and complexity. Assessment of individual and collective effectiveness in such a large-scale event is daunting but necessary for corrective feedback.

Summary and Recommendations

Nearly 3 decades of research and development on SP-based examinations makes it the most published innovation in medical education history. Students at all levels of education encounter a circuit of cases, sampled from the educational blueprint of a course or curriculum. Each sees the same patients as every other student. Predetermined checklists, rating scales, and questions posed to students apply the same criteria for evaluation for all students. Structured, simulation-based assessments of clinical performance have provided higher quality tools for objective, consistent evaluation of medical learners' clinical performance. Portrayal and marking of performance by SPs is consistent and accurate. Biases due to gender, ethic, and language differences have little to no impact on overall performance. Data from these examinations meet the need for more objective outcome data. However, that quality must be continually examined and challenged. Three areas are identified for advancement of SP-based examinations. Use of SP-completed checklists, although addressing logistical constraints early in development, are now inadequate for assessment of relatively advanced clinical performance. Post-clerkship assessments must include logic and organization of workups, degree of conceptual understanding, and situational awareness. Expected performance with a SP needs to incorporate the learner's use of information technology during the encounter to access diagnostic support and management options. Changes in the examination alone will not address the types of learning needed to perform at desirable levels. Some portion of clinical experiences must provide structured encounters with patients where students conduct

workups, are evaluated for cognitive approach, and technical quality of actions; then they must present a synthesized summary that will indicate their stage of understanding. Specific and supportive feedback must be given to each student who then must practice or read until deficits are overcome. Reassessment under formal conditions is then repeated. Current hospital and clinic experiences are necessary to prepare students for the tasks and responsibilities in the post-graduate component of training, but these experiences are simply inadequate to assure minimal competence.

Simulation-based assessments should go beyond the traditional clinical dyad to address multiple-person situations such as an individual practitioner's ability to handle multiple patients simultaneously or to see various family members over time. Theoretical frameworks for the professional development of expertise provide important information to guide new scenarios and instruments of assessment. Advanced constructs such situational awareness need to be assessed in scenarios at higher levels of education and clinical development.

Multiperson scenarios following those in aviation and anesthesiology should be developed for emergent and nonemergent situations. Review of data and videotape of the scenario provide a powerful learning opportunity. Discussion of what could be done differently can be followed by another run in the scenario. Structured assessment of teamwork should be extended beyond a formal testing situation. Simulated codes in the classroom, hospital-wide responses to catastrophic events, and even scenarios conducted in the community provide practice and reflection to improve performance and minimize errors. A very creative application of simulation-based assessment would be to evaluate a resident's conceptualization of "managing a ward service" as well as his or her performance in carrying out the management activities, monitoring his or her continuous situational awareness, and his or her adjustment in strategy as new issues develop. A similar approach could be developed for "managing the educational activities on a ward or in a clinic." Assessment of a senior resident's conceptualization of how to engineer learning for those at various levels of education who also are all part of the patient care team while being accountable to an attending physician (only one?!) who may or may not have the same situational awareness or team goals reflect the realities of residents. Reflection on strategy and alternative-solutions thinking should have immediate and beneficial effects. Finally, systems-based simulations of responses to attacks with weapons of mass destruction, for example, extended this assessment methodology beyond the health care system.

Research should continue to focus on the technical aspects of these new applications. Consistency of measurement, validity of interpretation, and generalization to similar scenarios are the traditional foci of measurement and evaluation specialists. Many of the recommended directions for advancing SP- and standardized scenario-based require broader, systems-level thinking. How do the assessments fit within the educational system? What kinds of evidence can we marshal that indicates they are having the intended impact on learning and performance? The most important aspect of these advancements is that require discussion and agreement on what is important and how to know when it is happening. Checklists, rating scales, oral examinations, written text, and objective question formats will follow easily. Reflection on the results of the new simulations by the developers will lead to questions like, "How do we know what we intended is actually happening?" Answers to those questions may be found in future research and in the literature. SP-based assessments have provided a key technology for improved evaluation of clinical performance. Creativity and research built on the accomplishments of the last 30 years will take assessment to the next level.

References

1. van der Vleuten CPM, and Swanson DB. Assessment of clinical skills with standardized patients: State of the art. *Teaching and Learning in Medicine* 1990;2:58–76.
2. Colliver JA, Williams RG. Technical issues: Test application. *Academic Medicine* 1993;68:454–60.
3. Petrusa ER. Clinical performance assessments. In *International handbook of research in medical education*. Dordrecht, The Netherlands: Kluwer Academic, 2002.
4. Furman G, Colliver JA, Galofre A. Effects of student gender and standardized-patient gender in a single case using a male and a female standardized patient. *Academic Medicine* 1993;68:301–3.
5. Colliver JA, Vu NV, Marcy ML, Travis TA, Robbs RS. Effects of examinee gender, standardized-patient gender, and their interaction on standardized patients' ratings of examinees' interpersonal and communication skills. *Academic Medicine* 1993;68:153–7.
6. Shatzer JH, Darosa D, Colliver JA, Barkmeier L. Station-length requirements for reliable performance-based examination scores. *Academic Medicine* 1993;68:224–9.
7. Shatzer JH, Wardrop JL, Williams RG, Hatch TF. The generalizability of performance on different-station-length standardized patient cases. *Teaching and Learning in Medicine* 1994;6:54–8.
8. Stillman P, Swanson D, Regan MB, et al. Assessment of clinical skills of residents utilizing standardized patients. *Annals of Internal Medicine* 1991;114:393–401.
9. Stillman PL, Regan MB, Haley HL, et al. Comparison of free-response and cued-response diagnosis scores in an evaluation of clinical competence utilizing standardized patients. *Academic Medicine* 1990;65:S27–8.
10. Harden RM, Stevenson M, Downie W W, Wilson GM. Assessment of clinical competence using objective structured examination. *British Medical Journal* 1975;1:447–51.
11. Cuschieri A, Gleeson FA, Harden RM, Wood RA. A new approach to a final examination in surgery. Use of the objective structured clinical examination. *Annals of the Royal College of Surgeons of England* 1979;61:400–5.

12. Harden RM, Gleeson FA. Assessment of clinical competence using an objective structured clinical examination (OSCE). *Medical Education* 1979;13:41–54.
13. Newble DI, Elmslie RG, Baxter A. A problem-based criterion-referenced examination of clinical competence. *Journal of Medical Education* 1978;53:720–6.
14. Newble DI, Elmslie RG. A new approach to the final examinations in medicine and surgery. *Lancet* 1981;2:517–8.
15. Newble DI, Hoare J, Sheldrake PF. The selection and training of examiners for clinical examinations. *Medical Education* 1980;14:345–9.
16. Petrusa ER, Guckian JC, Perkowski LC. A multiple station objective clinical evaluation. In *Proceedings of the ... Annual Conference on Research in Medical Education* (pp. 211–6). 1984.
17. Petrusa ER, Blackwell TA, Rogers LP, Saydjari C, Parcel S, Guckian JC. An objective measure of clinical performance. *American Journal of Medicine* 1987;83:34–42.
18. Stillman PL, Swanson DB, Smee S, et al. Assessing clinical skills of residents with standardized patients. *Annals of Internal Medicine* 1986;105:762–71.
19. Barrows HS, Williams RG, Moy RH. A comprehensive performance-based assessment of fourth-year students' clinical skills. *Journal of Medical Education* 1987;62:805–9.
20. Reznick RK, Blackmore D, Dauphinee WD, Rothman AI, Smee S. Large-scale high-stakes testing with an OSCE: Report from the Medical Council of Canada. *Academic Medicine* 1996;71:S19–21.
21. Conn HL Jr. Assessing the clinical skills of foreign medical graduates. *Journal of Medical Education* 1986;61:863–71.
22. Sutnick AI, Stillman PL, Norcini JJ, et al. ECFMG assessment of clinical competence of graduates of foreign medical schools. Educational Commission for Foreign Medical Graduates. *Journal of the American Medical Association* 1993;270:1041–5.
23. Online article. Retrieved from http://www.usmle.org/news/newscse.htm
24. Online article. Retrieved from http://www.nbme.org/examiner/fallwinter2002/news.htm
25. Kassebaum DG. The measurement of outcomes in the assessment of educational program effectiveness. *Academic Medicine* 1990;65:293–6.
26. Online article. Retrieved from http://www.aamc.org/publications/functionsandstructure.htm
27. Harden R. Personal communication, December 11, 2002.
28. Kern DE. *Curriculum development for medical education: A six-step approach.* Baltimore: Johns Hopkins University Press, 1998.
29. Norcini JJ, Shea JA. The reproducibility of standards over groups and occasions. *Applied Measurement in Education* 1992;5:63–72.
30. Whitcomb ME, Anderson MB. Transformation of medical students' education: Work in progress and continuing challenges. *Academic Medicine* 1999;74:1076–9.
31. Hunt CE, Kallenberg GA, Whitcomb ME. Trends in clinical education of medical students: Implications for pediatrics. *Archives of Pediatrics and Adolescent Medicine* 1999;153:297–302.
32. Gordon J, Hazlett C, Ten Cate O, et al. Strategic planning in medical education: Enhancing the learning environment for students in clinical settings. *Medical Education* 2000;34:841–50.
33. Halamek LP, Kaegi DM, Gaba DM, et al. Time for a new paradigm in pediatric medical education: Teaching neonatal resuscitation in a simulated delivery room environment. *Pediatrics* 2000;106:E45.
34. Smith PL, Ragan TJ. *Instructional design.* New York: Wiley, 1999.
35. Petrusa ER, et al. A multi-institutional trial of an objective structured clinical examination. *Teaching and Learning in Medicine* 1991;3:86–94.
36. Newble DI. Eight years' experience with a structured clinical examination. *Medical Education* 1988;22:200–4.
37. Vu NV, Barrows HS, Marcy ML, Verhulst SJ, Colliver JA, Travis T. Six years of comprehensive, clinical performance-based assessment using standardized patients at the Southern Illinois University School of Medicine. *Academic Medicine* 1992;67:43–50.
38. Anderson DC, Harris IB, Allen S, et al. Comparing students' feedback about clinical instruction with their performances. *Academic Medicine* 1991;66:29–34.
39. Ericsson, KA. *The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games.* 1996.
40. Schmidt HG, Boshuizen HP. On acquiring expertise in medicine. *Educational Psychology Review* 1993;5:205–21.
41. Poldre PA, Rothman AI, Cohen R, Dirks F, Ross JA. *Judgmental-empirical approach to standard setting for an OSCE.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA, April 1992.
42. De Gruijter DNM. Compromise models for establishing examination standards. *Journal of Educational Measurement* 1985;22:263–9.
43. Elstein AS, Shulman LS, Sprafka SA, et al. *Medical problem solving: An analysis of clinical reasoning.* Cambridge, MA: Harvard University Press, 1978.
44. Cohen R, Rothman AI, Poldre P, Ross J. Validity and generalizability of global ratings in an objective structured clinical examination. *Academic Medicine* 1991;66:545–8.
45. Swartz MH, Colliver JA, Bardes CL, Charon R, Fried ED, Moroff S. Global ratings of videotaped performance versus global ratings of actions recorded on checklists: A criterion for performance assessment with standardized patients. *Academic Medicine* 1999;74:1028–32.
46. Hodges B, Regehr G, McNaughton N, Tiberius R, Hanson M. OSCE checklists do not capture increasing levels of expertise. *Academic Medicine* 1999;74:1129–34.
47. Bordage G, Lemieux M. Semantic structures of diagnostic thinking of experts and novices. *Academic Medicine* 1991;66:S70–2.
48. Bordage G. Elaborated knowledge: A key to successful diagnostic thinking. *Academic Medicine*. 1994;69:883–5.
49. Bordage G, Connell KJ, Chang RW, Gecht MR, Sinacore JM. Assessing the semantic content of clinical case presentations: Studies of reliability and concurrent validity. *Academic Medicine* 1997;72:S37–9.
50. Bordage G. Why did I miss the diagnosis? Some cognitive explanations and educational implications. *Academic Medicine* 1999;74:S138–43.
50. Dreyfus HL, Dreyfus SE. *Mind over machine: The power of human intuition and expertise in the area of the computer.* New York: Free Press, 1986.
51. Eruat M. *Assessment of competence in the professions.* Bristoal, PA: Falmer, 1993.
52. Eraut M. *Developing professional knowledge and competence.* Bristol, PA: Falmer, 1994.
53. Endlsey MR, Bolsted CA . Individual differences in pilot situation awareness. *International Journal of Aviation Psychology* 1994;4:241–64.
54. Gaba DM, Howard SK, Small SD. Situation awareness in anesthesiology. *Human Factors* 1995;37:20–31.
55. Artman H. Situation awareness and co-operation within and between hierarchical units in dynamic decision-making. *Ergonomics* 1999;42:1404–17.
56. Waag WL, Houck MR. Tools for assessing situational awareness in an operational fighter environment. *Aviation Space and Environmental Medicine* 1994;65:A13–9.

57. Paris CR, Salas E, Cannon-Bowers JA. Teamwork in multi-person systems: A review and analysis. *Ergonomics* 2000;43:1052–75.
58. Holzman RS, Cooper JB, Gaba DM, Philip JH, Small SD, Feinstein D. Anesthesia crisis resource management: Real-life simulation training in operating room crises. *Journal of Clinical Anesthesia* 1995;7:675–87.
59. Gaba DM, Howard SK, Flanagan B, Smith BE, Fish KJ, Botney R. Assessment of clinical performance during simulated crises using both technical and behavioral ratings. *Anesthesiology* 1998;89:8–18.
60. Spath P. *Error reduction in health care: A systems approach to improving patient safety.* San Francisco: Jossey-Bass, 2000.
61. Clay MC, Lane H, Willis SE, Peal M, Chakravarthi S, Poehlman G. Using a standardized family to teach clinical skills to medical students. *Teaching and Learning in Medicine* 2000;12:145–9.
62. Harward D. Personal communication, September 19, 2000.
63. Lesky LG, Wilkerson L. Using "standardized students" to teach a learner-centered approach to ambulatory precepting. *Academic Medicine* 1994;69:955–7.
65. Howard SK, Gaba DM, Fish KJ, Yang G, Sarnquist FH. Anesthesia crisis resource management training: Teaching anesthesiologists to handle critical incidents. *Aviation Space and Environmental Medicine* 1992;63:763–70.
64. Rethans JJ, Sturmans F, van der Vleuten CPM. A method for introducing standardized (simulated) patients into general practice consultations. *British Journal of General Practice* 1991;41:94–6.
65. Rethans JJ, Sturmans F, Drop R, van der Vleuten CPM. Assessment of the performance of general practitioners by the use of standardized (simulated) patients. *British Journal of General Practice* 1991;41:97–9.
66. Gordon M. Personal communication, October 20, 2002.

Received final 20 December 2002

APPENDIX
All Actions for Hernandez Case
(Critical Actions are in boldface)

History (SP marks that the student determined …)

1. **How long I have had this cough?**
2. How my appetite has been while I've had this cough?
3. If the cough brings up any stuff?
4. If I noticed any blood in it?
5. If I had pain in my chest or elsewhere when I am coughing?
6. What kind of work I do?
7. **If I wake up in the night sweating?**
8. If anyone at work or at home have a cough like mine?
9. Who I am living with?
10. If I get short of breath more easily since I've had this cough?
11. If I drink alcohol or use street (recreational; illegal) drugs?
12. How much I drink?
13. If I have ever passed out after drinking?
14. How long ago was that?
15. How is my parents' health?
16. If I know of any family members or close friends who have had TB?
17. Where I am from (country of origin)?
18. How long I have been in the US?
19. If I have I ever been tested for AIDS or HIV?

Physical Exam

20. Tapped his/her finger on your chest and back
21. Placed both hands on your back and had you breathe normally
22. With hands still on your back, asked you to say "99" or "moon"
23. **Listened with the stethoscope to both sides of your chest in the front**
24. **Listened with the stethoscope to your back on both sides**
25. While listening to your back, asked you to say "eeeeee"
26. **Requested you cover your mouth when you cough and/or gave you a paper mask to wear**
27. Performed all or part of the physical examination through the gown

Relationship (Scale: 1 [*poor*], 2 [*fair*], 3 [*good*], 4 [*very good*], 5 [*excellent*])

With this student doctor to what degree did you feel:

1. **Respected**[a]
2. **Comfortable/at ease**[b]
3. **Understood**

Communication (Scale: 1 [poor], 2 [fair], 3 [good], 4 [very good], 5 [excellent])

How was the student doctor you just saw at:

4. **Greeting you warmly, being friendly, never crabby or rude**[b]
5. **Treating you like you're on the same level, never "talking down" to you or treating you like a child**
6. **Letting you tell your story, listening carefully, asking thoughtful questions, not interrupting**
7. **Showing interest in you as a person, not acting bored or ignoring what you have to say**
8. **Encouraging you to ask questions, answering them clearly, never avoiding your questions or lecturing you**
9. **Using easily understood words and explaining any technical or medical terms in plain language**

Assessment and Plan—Based on the Information You Have Gathered, List the Top 5 Diffenential Diagnoses You Are Considering at This Time

Diagnosis #1: TB as one of top 3
Diagnosis #2:
Diagnosis #3: Lung Cancer in list
Diagnosis #4:
Diagnosis #5:

Select any Diagnostic Studies You Would Order at This Time: (From List of 98 Diagnostic Studies)
Select: Chest X_ray

Note: Critical actions are in bold. SP = standardized patient; TB = tuberculosis.
[a]All three items must be rated 2 or higher. [b]No more than one item rated 2 or less.