

2001

Medical Education again Provides a Model for Law Schools: The Standardized Patient Becomes the Standard Client

Lawrence Grosberg

New York Law School, lawrence.grosberg@nyls.edu

Follow this and additional works at: http://digitalcommons.nyls.edu/fac_articles_chapters

Recommended Citation

51 J. Legal Educ. 212 (2001)

This Article is brought to you for free and open access by the Faculty Scholarship at DigitalCommons@NYLS. It has been accepted for inclusion in Articles & Chapters by an authorized administrator of DigitalCommons@NYLS.

Medical Education Again Provides a Model for Law Schools: The Standardized Patient Becomes the Standardized Client

Lawrence M. Grosberg

A teaching technique used in medical schools has proved to be quite valuable in a law school setting as well: playing the part of a Standardized Patient, an actor or other layperson is interviewed and examined by a medical student. Afterwards the SP provides written feedback to the student, using an evaluation checklist prepared by the medical faculty. The form assesses the student's clinical performance. This teaching tool has been a part of medical education, especially in the third and fourth years, for more than twenty years. Its use is so widespread that it is now one of an array of testing methods on the medical licensing exams.

Notwithstanding its extensive use in the medical world, the SP concept has not been copied in law schools, at least not to my knowledge and not until the experiments described in this essay. The striking similarities between certain aspects of clinical education in medical and law schools suggested to me that we could learn something from our medical colleagues.¹ In both professional schools a primary pedagogical objective is to teach students how to apply their medical or legal knowledge in context, with the specific situation of each patient or client in mind. I went forward with a pilot project using the SP concept because it seemed promising as an effective and cost-efficient way to provide skills training to large numbers of law students. I also thought it could be an excellent vehicle to supplement traditional methods of doctrinal teaching. And it offered a potential vehicle for evaluating law students' interviewing

Lawrence M. Grosberg is a professor at the New York Law School.

I thank Ellen Ryerson and Stephen Ellmann for their very useful comments on earlier drafts of this essay. I am also grateful to Dr. Mark Swartz and Devra Cohen of the Mt. Sinai Medical School for graciously allowing me to observe and discuss with them their use of the standardized patient. Finally, I want to thank Harry Wellington, Ellen Ryerson, Richard Matasar, and New York Law School for their support of the pilot project discussed in this essay and other continuing efforts at improving legal education.

1. Both physicians and lawyers must interview and counsel, often about very personal and difficult matters. Both must first learn details from the interviewee and then must communicate frequently complex information in a clear, comprehensive, humane, and non-condescending manner.

and counseling skills. For this pilot, the role-playing layperson became a client or a witness who is interviewed by a law student. That Standardized Client then provides written feedback to the student on a checklist and rating form prepared by the law faculty. I describe the very positive results below.

In the context of shrinking law school resources, expansion of intensive one-on-one clinical instruction faces considerable resistance and competition. We have to ask whether there are alternative ways to teach and evaluate clinical lawyering skills that are effective but less expensive. The standardized client seems to be one answer. In this instance, the economics of pedagogy could be determinative. At New York Law School the SC has become an integral part of our first-year curriculum. It also seems fair to say that the SC could be a successful complement to our live-client clinics and many other experiential teaching methods that we and others use.

In this essay I first describe how medical schools have used the standardized patient (part I). Next I summarize the extensive empirical research that supports using the SP as a method for teaching and evaluating medical clinical skills (II). I then describe our standardized client project at NYLS (III) and go on to explore the potential value of the SC in legal education (IV) and our need for an efficient administrative structure to use SCs and collect empirical data (V). If law schools were to adopt the model I posit—one whose premise calls for establishing a consortium among small groups of law schools—we might find that the value of the SC and the educational gains to each school could be far greater than the sum of all the schools' financial contributions.² Finally I discuss briefly the utility of the SC concept as an additional high-stakes testing device (VI). Theoretically at least, the inclusion of the SC on bar exams could bring the ultimate bar admission hurdle in line with a more holistic notion of what a competent lawyer must know before being licensed to practice law. But that day is a long way off.

I. The Standardized Patient

A. *The Genesis of the SP in Medical Education*

Clinical education—learning by doing and learning by observing doctors and teachers practicing medicine, typically in a university or teaching hospital—has long been an integral part of the medical school curriculum.³ In this sense, medical education has been far ahead of legal education in preparing

2. Of course there are limits as to the propriety of allowing new revenue sources or cost savings to be the primary motivation or justification for a particular pedagogy. Richard A. Matasar has promoted the fee-for-service model of clinical education as one way to address the resource issue in *The MacCrate Report from the Dean's Perspective*, 1 *Clinical L. Rev.* 457 (1994). But see Martin Guggenheim, *Fee-Generating Clinics: Can We Bear the Costs?* 1 *Clinical L. Rev.* 677 (1995); Lisa A. Lerman, *Fee-for-Service Clinical Teaching: Slipping Toward Commercialism*, 1 *Clinical L. Rev.* 685 (1995).
3. Experiential learning continues in postgraduate residencies as well. From 1915 until 1963 the National Board of Medical Examiners included as part of the process of licensure an interview/examination of an actual patient, observed by an evaluator. See Daniel J. Klass, "High Stakes" Testing of Medical Students Using Standardized Patients, 6 *Teaching & Learning in Medicine* 28, 29 (1994).

its graduates to practice the profession. But while clinical legal education has increased in recent years, external pressures in the medical profession have led to a lessening of its focus on physician-taught individualized clinical learning.⁴ Further, the use of direct physician-observed clinical evaluations of medical license applicants has been called into question; many believe that this method of assessment is inconsistent, unreliable, and ultimately unfair.⁵ These were not issues of personalized bias or unfairness by physician evaluators. Rather, the customary standards for establishing the reliability and validity of a testing device that is used in high-stakes testing (such as a test for a professional license) were not being satisfied.⁶ All of these pressures—on medical school faculty and on those licensing evaluators—have led to wider acceptance of the standardized patient.⁷

Medical educators began using actors to simulate patients more than thirty years ago. The actor would be carefully trained to assume the role of a patient, incorporating a general biographical profile of the patient with the patient's emotional and psychological traits as well as a wide variety of physical symptoms. The standardized patient could even be an actual patient with a disease of one sort or another who has real physical symptoms that an actor would not be able to simulate.⁸ A person with a real illness would receive the same training as an actor. The term *standardized patient*, rather than *simulated patient*, came to be used for the person performing the dual acting/assessing function because it is a more comprehensive term, including both simulated patients and actual patients "who have been carefully coached to present their own illnesses in a standardized unvarying way."⁹ The training of SPs is quite thorough; it aims to standardize both their role-playing and their evaluating. As for the evaluation checklists, the goal is accuracy as well as uniformity in the

4. See Consensus Statement of the Researchers in Clinical Skills Assessment (RCSA) on the Use of Standardized Patients to Evaluate Clinical Skills, Appendix 1 to Special Issue: Proceedings of the AAMS's Consensus Conference on the Use of Standardized Patients in the Teaching and Evaluation of Clinical Skills, 68 *Acad. Med.* 475 (1993) [hereinafter Consensus Statement]. One medical professor is quoted, in support of her SP research and recommendation that SPs be used in medical school: "[M]edical students are rarely if ever observed [during the four years of medical school] by faculty physicians in taking a history and performing a physical examination . . ." Mark Swartz & Jerry Coliver, Further Discussion of SP Checklists and Videotaped Performances, 75 *Acad. Med.* 315, 317 (2000).
5. These medical license evaluations took place in hospitals; they were hostage to the type of patients that happened to be in the hospital at the time and to the scheduling of the physician-evaluators. It became increasingly difficult to ensure uniformity and fairness.
6. For example, an evaluator's assessment of a licensure applicant interviewing an appendicitis victim is not comparable, in a statistical psychometric sense, with that evaluator's assessment of an applicant interviewing a skin cancer patient. Similarly, if different faculty evaluated different applicants, the ratings could vary substantially.
7. For a comprehensive history of the standardized patient, see Howard S. Barrows, An Overview of the Uses of Standardized Patients for Teaching and Evaluating Clinical Skills, 68 *Acad. Med.* 443 (1993).
8. For example, the SP might be a person with a heart murmur, and the medical student would, among other things, have to discover the heart murmur. Such SPs were sometimes called patient instructors.
9. Barrows, *supra* note 7, at 444.

completion of the faculty-drafted forms. If medical students cannot be observed taking medical histories by their faculty, they can still conduct such an interview and receive some feedback from the SP. This continues to be the educational rationale for use of the method.

There are several key features in the design and use of the SP exercises. First, the case history of the patient is carefully designed by the medical faculty and then pretested. Second, a similar group of medical faculty, typically a committee, drafts and refines an evaluation checklist that addresses the components of clinical skills that the group deems necessary to a competently performed interview with the patient.¹⁰ Depending on the nature of the interview and the pedagogical objectives, there may also be a questionnaire for the student to complete that covers both substantive medical questions (a diagnosis and prescribed treatment, for example) and a self-critique of interpersonal skills used in the session.¹¹

Typically the SP evaluation checklist calls on the SP to record in yes/no fashion whether certain information was elicited (e.g., a chest pain symptom) or whether a particular action or behavior took place. The SP is expected to “document actions performed by students on history taking and physical examination and behavior related to interpersonal and communication skills.”¹² Occasionally an SP may also be asked for a more global rating—an evaluation, for example, on a 1-to-4 scale (poor to excellent) of the student’s communication skills.¹³ Generally, however, medical educators and licensing authorities have been reluctant to ask laypersons to give an overall grade—an impressionistic evaluation—for students’ performances.

B. The Current Uses of SPs

SPs are now used primarily for two different purposes and are beginning to be used for a third purpose. These uses exactly parallel the potential utility of the concept in legal education. In one current use, the SP exercise is graded and included among the four-year requirements for an M.D. degree. This graded assessment might involve either a traditional grade or a pass/fail standard, but in either case the student must demonstrate clinical or commu-

10. To assist the reader without burdening this essay, I have established a Web site containing various supporting materials to which I refer. A sample standardized patient checklist is included as Appendix A at <<http://www.nyls.edu/grosberg>>. The documents are posted in PDF format, which requires that you have the Adobe Acrobat Reader installed on your computer. Instructions on downloading this free program also are available on my Web site.
11. There is ongoing research on the accuracy of these student self assessments. James T. Fitzgerald et al., *The Influence of Task Formats on the Accuracy of Medical Students’ Self-Assessments*, 75 *Acad. Med.* 737 (2000).
12. Mark H. Swartz & Jerry Colliver, *Using Standardized Patients for Assessing Clinical Performance—An Overview*, 63 *Mount Sinai J. Med.* 241, 241 (1996).
13. For the most part SPs were asked for global ratings only for the limited purpose of collecting empirical data and not as part of the customary feedback that they provided to students. See D. S. Cohen et al., *A Large-Scale Study of the Reliabilities of Checklist Scores and Ratings of Interpersonal and Communication Skills Evaluated on a Standardized-Patients Examination*, 1 *Advances in Health Sci. Educ.* 209 (1997).

nication skills at a minimal level of competence to satisfy the requirement.¹⁴ In the context of four years of the typical medical school, such graded exercises are a very small part of the overall evaluation of students, given the great number of other subjects taught (customarily in lecture format) and graded (usually by written objective exams). Nevertheless, an increasing number of schools require SP assessments. Typically the scoring on these exams is based on the percentage of *yes* responses.

A second use of SPs is for learning purposes only: the student may participate in one or more SP exercises, but is not required to achieve a minimal competence level.¹⁵ Most of the growth in the actual use of SPs is in this category of constructive feedback. Both the graded and the learning-only uses of the SP reflect the unanimous belief among medical educators that clinical skills are of central importance in a doctor's education, and their recognition that experiential learning is essential to the development of clinical skills. The only question is how to ensure that students' clinical skills are adequately taught and evaluated. Ostensibly the medical schools acknowledge that at least some of the assessments should be done through direct observation by medical faculty. The reality, however, is something else.¹⁶ The need to remedy the absence of individual feedback from faculty has been among the forces behind the increasing use of the SP.

A third use of SPs that has engendered considerable controversy is as part of the physicians' licensing examination. While there is some empirical support for requiring satisfactory completion of an SP exercise for entry to the medical profession, it is not yet included on the standard licensing exam. There remain questions about its validity, and there are practical and logistical obstacles to its inclusion. But the SP is already included on the licensing examination taken by foreign-educated physician applicants.¹⁷ A key distinguishing rationale is that the technique tests, among other things, the

14. See, e.g., Paula Stillman, Guidelines for the Use of Standardized Patients, Paper Prepared for the 1996 Annual Meeting of the American College of Physicians, April 19, 1996 (on file with author). The graded SP could be further broken down into a "medical school progress test" or an "end-of-training achievement test."
15. I observed such an exercise at Mt. Sinai Medical School—one of several for third-year students. During the year the students rotate through the medical specialties. Each rotation brings the student into the teaching hospital to observe residents and attending physicians examining and interviewing patients in their specialty. In the course of each rotation, each student interviews and examines a SP in that specialty. My observations of the SP exercise are described in a narrative memo included as Appendix B at my Web site, *supra* note 10.
16. In a recent survey of medical schools, 98 percent of the schools reported that they use faculty to assess student clinical skills through direct observations. In contrast, a similar survey of medical school graduates disclosed that only 73 percent of them reported having their clinical skills evaluated by faculty observation. See The Role of Faculty Observation in Assessing Students' Clinical Skills, *Contemp. Issues Med. Educ.*, Oct. 1997, at 1.
17. These exams are overseen by the Educational Commission for Foreign Medical Graduates. The ECFMG works with the National Board of Medical Examiners, which administers the licensing exams for U.S. applicants, but it is an independent organization. See Gerald Whelan, High-Stakes Medical Performance Testing: The Clinical Skills Assessment Program, 283 *JAMA* 1748 (2000); Amitai Ziv et al., Lessons Learned from Six Years of International Administrations of the ECFMG's SP-Based Clinical Skills Assessment, 73 *Acad. Med.* 84 (1998).

applicant's ability to communicate in English.¹⁸ Such a test is less necessary for an applicant who is a native speaker.

In any event, when considering the use of standardized patients or standardized clients, it is important to be clear as to these significantly different purposes of the method.

II. The Importance of Empirical Data

The different uses of the SP illuminate the critical role of the empirical research that addresses the effectiveness of the SP. In turn, what must we do if we are to contemplate using the legal analog, the standardized client? Looking at the vast empirical work our medical colleagues have already completed is invaluable. The different functions for which the SP or SC may be used clearly demand supporting data of appropriately varying quality.

When a student receives a teacher's feedback on a performance in a classroom setting (e.g., an oral colloquy on the symptoms of a disease in medical school or a Supreme Court opinion in law school) but the evaluation has no bearing on the student's grade in that class, there is minimal concern about the validity of the evaluation or compliance with standard testing procedures. The assessment is solely for the learning benefit of that student, or for the students who are listening. At the other end of the spectrum is someone who is being evaluated for a license to practice medicine or law. That person will be very much concerned with whether the evaluation process is a fair, reliable, and valid measure of the applicant's ability to be a physician or a lawyer. In between, there might be a low-stakes exam that determines part or all of a grade in a single class required for graduation. In that situation, the student would like the assessment tool to be as fair and accurate as possible, but will not be as much concerned as in the case of a licensing exam. Psychometric testing standards are less critical.

With respect to the SP or the SC as an evaluative device, there is an added challenge. How would a medical student or a law student view an assessment by a layperson, even one offered only for constructive learning purposes? Would it be accepted as a valid evaluation of professional competence? Could a physicist assess a social worker's skills as a counselor? It seems counterintuitive. If only to motivate the student to take seriously the SP or SC assessment tool, it would seem essential that it be accepted as a reasonably fair and accurate method of feedback. That is one of the purposes of the empirical studies.

What have our medical colleagues accomplished? Over the last twenty years, they have completed an extraordinary amount of empirical research on SPs, usually conducted by psychometricians and medical faculty. In terms of substantiating the value of the SP for learning purposes, the data are positive and "reassuring."¹⁹ First, the SPs were found to be realistic. In tests when SPs went unannounced to physicians' offices, the doctors could not tell the SP

18. It is worth noting, parenthetically, that Mt. Sinai Medical School has joined forces with Kaplan Educational Services to conduct a prep course for the SP exam for graduates of foreign medical schools.

19. Consensus Statement, *supra* note 4, at 476.

from the real patient.²⁰ There is evidence that students seriously prepared for the SP exercises and then conducted them in the same manner as with real patients.²¹ Empirical studies also showed that the SPs accurately portrayed the profiled patient they were supposed to be playing.²² A study also showed that the SPs were reasonably accurate (82%) in recording what actually happened at the meeting—what the medical student said and did.²³ On the issue of reliability—in very simple terms, the consistency and stability of the measurements over repeated exercises—the data on SPs also are reasonably positive.²⁴

The most fundamental question is whether the SP validly assesses the student's clinical skills. In simple terms, validity means the test in fact evaluates what it intends to evaluate. To establish the validity of the SP, researchers must show that the SP quantitative scores really measure the "construct of clinical competence."²⁵ Variance in SP scores must correlate to variations in the level of actual competence of the test takers. To verify the correlation, researchers must have a measure that closely approximates actual clinical competence; they call it the gold-standard criterion. They have concluded that there is only one, and that is the global rating by faculty-physicians. These are individual subjective evaluations by the faculty as to the clinical competence of each student.

A major study was recently completed that compared SP scores with faculty global ratings.²⁶ The study examined the videotaped SP exercises of fifty medical students (randomly selected from hundreds) who participated in the SP program at Mt. Sinai Medical School. The students came from eight medical schools that participate in the Mt. Sinai SP consortium (discussed below). Each student conducted seven SP sessions; five faculty then viewed all of the videotapes and rated each student on each exercise and then each student overall, for all seven sessions. The researchers then compared the SP scores with the faculty global ratings.²⁷ They concluded that the correlations between the SP checklists and the teachers' ratings were not as close as they had expected. The teachers' ratings for "clinical competence" (diagnostic examination actions) randomly selected were higher than those of the SPs,

20. Swartz & Colliver, *supra* note 12, at 244.

21. See generally Geoffrey R. Norman et al., *Simulated Patients*, in *Assessing Clinical Competence* 224 (New York, 1985). See also my narrative memo, Appendix B at my Web site, *supra* note 10.

22. Swartz & Colliver, *supra* note 12, at 244.

23. *Id.*

24. *Id.* at 245.

25. *Id.* at 247.

26. See M. H. Swartz et al., *Global Ratings of Videotaped Performance Versus Global Ratings of Actions Recorded on Checklists: A Criterion for Performance Assessment with Standardized Patients*, 74 *Acad. Med.* 1028 (1999).

27. The typical SP checklist would have three groupings of yes/no questions. The first would be a list of facts that the student did or did not ascertain; the second would be a list of examination steps the student did or did not do; and the third would be a list of communication skills that the student did or did not demonstrate. The number of yes and no answers would be computed to come up with the SP scores.

whereas faculty ratings for “interpersonal and communication skills” were lower.²⁸ So the Mt. Sinai research team questions whether a SP assessment, at this point, may be completely interchangeable with a teacher’s rating of an observed performance and argues, at least for the time being, that SP use on the high-stakes licensing exams should not be encouraged.²⁹

On a less scientific level, the use of SPs has received a very positive response in the popular press and by medical students. One of the more public comments was a recent *Newsweek* report on medical schools applauding SP use as reflective of the need for physicians to improve their “communication skills.”³⁰ There also has been a steady stream of articles very favorably reporting on medical schools’ incorporation of the SP into their curriculum.³¹ Finally, it should be noted that the success of the medical educators has led other professions to use the same SP technique.³²

The lesson for legal education seems clear. If we want to emulate our medical colleagues, we must take much more seriously the need to verify our instincts regarding various modes of evaluating lawyering performances. We must carefully compare, for example, the accuracy of nonlawyer evaluations with those of a law teacher. Both the law students and the bar, as well as the public, must have sufficient confidence in the method to accept it as fair and valid, whether for the purposes of constructive learning and feedback or for more consequential testing. To that end, I have undertaken a series of experiments aimed at evaluating the use of the SC. Much more must be done.

III. The NYLS Standardized Client Pilot Project

Simulation has been used in legal education for some time—both prearranged role-playing by actors, students, or others assuming (typically) the roles of clients or witnesses and extemporaneous role-playing in the classroom, with teacher or students or both assuming a role. These teaching techniques were first used in upper-level clinics and skills courses, particularly

28. One possible explanation is that certain of the items on the checklist may be less or more important to the physician-evaluators, and the lists should be refined to more closely approximate the actual components of the physicians’ assessments. See Swartz et al., *supra* note 26, at 1030–31.
29. This cautiousness is in accord with other SP supporters who say that messianic or missionary zeal for the SP could backfire if changes were pushed too quickly. See George E. Miller, Proceedings of the AAMC’s Consensus Conference on the Use of the Standardized Patients in the Teaching and Evaluation of Clinical Skills, Conference Summary, 68 Acad. Med. 471, 472 (1993).
30. Bronwyn Fryer, Brave New World, How to Get into Graduate School, 1998 ed., *Newsweek*, 1997, at 36.
31. See, e.g., Ronald Rutti, Faking Illness for Good Health: Patient Pretenders Part of Lessons at Medical School, Clev. Plain Dealer, Apr. 26, 1999, at B1.
32. See, e.g., J. Hampl et al., Using Standardized Patients to Train and Evaluate Dietetics Students, 99 J. Amer. Dietetic Ass’n 1094 (1999); M. H. Mariano et al., The Use of Standardized Patients to Supplement Clinical Education of Physical Therapy Educational Program, 79 Physical Therapy S47 (1999).

in trial advocacy, but by now they have expanded into the traditional classroom as well.³³ While the role-player client or witness occasionally gives feedback, it generally is an informal affair that depends on the time and availability of the role-player or the student. And when it does occur, the feedback is instinctual for the most part, typically of widely disparate quality among the role-players and not designed in any systematic way to achieve any specific skills-learning objectives.³⁴

Using a well-trained layperson as a role-player whose specific responsibilities include giving explicit written feedback to the student is, I believe, a new feature in legal education. What it adds to our simulation techniques is the notion that the students will have a much more standardized common experience, both in interacting with the client or the witness and in getting relatively uniform evaluation, all without the direct intervention of a faculty member. To test the idea, I began an experiment, based on the successes of the SP in medical schools, that used the same concept in my law school.³⁵ I undertook a three-stage pilot project using a single standardized client exercise, first in an upper-level simulation skills course (43 students); second, in one section of a first-year required introductory lawyering-process course (110 students); and third, in all four sections (450 students) of that same lawyering course.³⁶

A. First Stage: The Upper-Level Experiment

In a class on negotiating and counseling (43 students), I modified an exercise I had been using to include a standardized client who would be counseled by the students.³⁷ The counseling exercise was the first of two graded simulations in which the student participated; the second was a negotiation the student conducted on behalf of the client previously counseled.

33. See Lawrence Grosberg, *Should We Test for Interpersonal Lawyering Skills?* 2 *Clinical L. Rev.* 349, 353 (1996).

34. One possible exception occurs when students assume prearranged client roles and are assigned explicitly the additional task of completing a structural critique of the fellow student performing the lawyer role. But my experience is that this method yields mixed results at best. Because students do not view their evaluation role (even when assigned that task) as their central responsibility (as a SC does), the quality of student feedback varies greatly.

35. The interpersonal tasks of the physician and the lawyer are quite similar: interviewing, fact-finding, and communicating medical or legal principles or concepts in a clear, comprehensive, empathetic, and noncondescending manner. Cf. Linda F. Smith, *Medical Paradigms for Counseling: Giving Clients Bad News*, 4 *Clinical L. Rev.* 391 (1998).

36. While I was attempting to implement and administer a SC program, I was also trying to collect and compile data that might support the validity of the SC. Doing this for the first time, I struggled in the effort to accomplish both objectives. Upon completion of the three-stage pilot project, and after reviewing my internal reports (the major contents of which are reflected in this essay), the law school administration regularized the inclusion of the costs of the exercise in our annual budget. We are about to embark on a fourth stage in which we will triple our program: each of our 450 first-year students will do three SC exercises.

37. The client was the CEO of a drug company accused of firing an executive for blowing the whistle on some questionable decisions relating to the disclosure of drug research. The putative plaintiff had not yet filed the complaint, and the student had to counsel the client on what to do. The student and the client had a maximum of 45 minutes for the counseling session.

In preparing the SC part of the exercise, I first drafted an evaluation checklist for the SCs to complete after the counseling session. Following the medical school model, I met with the group of teachers with whom I teach the negotiating and counseling course, and we collectively drafted and revised the checklist.³⁸ I then hired actors to play the client and trained them both to play the role and, more important, to evaluate the students by completing the checklist. The latter required several sessions during which the actors viewed videotapes of students doing the counseling, individually completed the checklists, and then discussed their evaluation decisions with one of the teachers of the course.

My primary objective at that initial stage of the pilot was to assess the actors' ability to evaluate the students' performances. I felt confident about the role-playing part, although even that required some changes from our prior preparation practices. For example, I trained the three actors together, and we collectively analyzed their videotaped rehearsals of the role. My prior preparation of role-players had been much more informal and ad hoc; usually I was alone with each individual role-player for a very short time. If the SCs might later do some actual grading,³⁹ the experience (both the actual counseling and the feedback or grading) for each student has to be standardized; even the training for the role-playing required a focused effort to ensure that each SC acted essentially the same way. While we allowed for some personality differences among the actors, we tried to keep them within fairly narrow limits.

Each student was videotaped while counseling the actor-client, so the teachers could view the tape, grade the performance in their customary manner, and then complete the checklist. This approximated the way in which each actor completed an identical checklist just after the counseling session.⁴⁰ I then compared the checklists completed by the actors with those completed by the teachers for the same students.⁴¹ We compared the quantitative subtotals of all questions within each part, the individual subtotals of the

38. A copy of the checklist for that SC exercise is included as Appendix C at my Web site, *supra* note 10. The evaluation form had four parts, setting forth the criteria for assessing: (1) general communication skills; (2) the organization of the meeting; (3) the clarity and thoroughness of the session; and (4) the appropriate decision-making role for the lawyer. Each of the parts had a series of items. The SC rated each item 0, 1, or 2 and also gave an overall rating for each part. There also was a section for the SC to include other comments.

39. At this experimental stage the SC assessments played no role in the actual grading of the students.

40. Since the teachers completed the checklists after they had recorded their grades for each student, the SC aspect of the exercise had no impact on the students' grades. It is worth noting that completing a checklist after a live performance is not necessarily the same as doing so after viewing a videotape. With respect to SPs, this was pointed out in a recent study comparing the assessments of SPs and medical faculty. See André De Champlain et al., *Standardized Patients' Accuracy in Recording Examinees' Behaviors Using Checklists*, 72 *Acad. Med. Supp.* S85 (1997).

41. Unlike the typical SP yes/no checklists, the ones I used in this first stage asked for ratings: 0, 1, or 2. Similarly, I asked the SCs and the teachers for a global overall rating in each of the four parts of the checklist. I could then compile quantitative totals for each part of the checklist.

overall assessments for each category, and the totals of all overall assessments. While the evaluations of the SCs and the teachers were by no means identical, the results showed a positive correlation between the teachers' and the actors' evaluations of the student performances.⁴²

There were severe resource limitations in our empirical analysis.⁴³ We had no help from a psychometrician in planning or analysis. Unlike the Mt. Sinai programs where each student interviewed as many as seven SPs, my project had each law student seeing only one client. We had no crosschecks of the student performances with different clients (to confirm consistency in a student's performance in different contexts), a necessary step to demonstrate validity and reliability in the assessments. The numbers of students, actors, and teachers were relatively small, so that it was impossible to draw any conclusions based on statistically significant figures.

I examined the degree to which the evaluations of the actors and the teachers correlated. To an even more limited extent, I could see if my own assessments of students and those of the teachers and actors were similar. (I also viewed the tapes and completed an evaluation checklist.⁴⁴) The results, while limited, were encouraging.⁴⁵ They suggested that on key aspects of the evaluations, there was some consistency between the assessments of the actors and those of the teachers.

The most significant positive association was the correlation coefficient of the total of overall assessments—0.50.⁴⁶ What this seems to say is that there was at least some consistency in the overall evaluative conclusions of the actors and the teachers. Also, the correlation coefficient representing the totals of all of the scores for all of the items in each of the four categories was 0.47.⁴⁷ Again, while none of these figures demonstrates the statistical validity or reliability of

42. We also compared the actors' assessments of particular items in a category with the overall assessment for that category to see whether there were some items that had a stronger correlation with the overall assessment scores than others. Even with this limited number of students and with only one exercise, there were still other comparisons that we did not examine. For example, we did not compare the scoring of individual actors or individual teachers even though a visual review would suggest that there were differences between them in their scoring.

43. These limitations underscore the importance of conducting more sophisticated studies if the SC is ever to approach the widespread use of the SP.

44. I refrain from using the term *subjective assessment*. All performance evaluations—whether of a traditional essay exam answer, an advocate's appellate brief, a counseling performance, or a jury summation—involve the individualized personal assessment of the evaluator. But in each instance the evaluation should be grounded in generally accepted criteria.

45. I thank Brian Lansbury for his excellent assistance in analyzing the data while he was a third-year student at NYLS.

46. The correlation coefficient is a quantitative representation of the association between two variables. It is a measure of linear association or clustering along a line. See James Brook, *A Lawyer's Guide to Probability and Statistics*, 215–19 (Toronto, 1990).

47. The first comparison was the total of each of the global assessments for the four categories on the evaluation form. The second comparison was the total from all of the questions on the form. Two scatter grams depicting the correlation coefficients for these two comparisons are included as Appendix D at my Web site, *supra* note 10.

this method of evaluating students' performances, they do show some consistency, enough to justify the use of SCs for ungraded exercises and perhaps enough for low-stakes grading.⁴⁸ To reiterate, and to keep these results in perspective, such an empirical analysis of any law school testing is rare. One might speculate about what the results would be if we were to evaluate the typical law school exam against the same rigorous psychometric standards.

There are many similarities between legal and medical education in the structure and purpose of their use of trained actors. Having observed the feedback techniques in both fields, I concluded from my first experiment that individual feedback from trained nonlawyers (standardized clients) can be a valuable learning method if used in conjunction with teaching that places the experiential exercises in context. Classroom discussion of the students' common experience with the same exercise enhanced its learning value. I decided to continue our experimental use of these evaluation methods, building on the experience of our medical school colleagues. In the initial stage I had purposely used a counseling exercise that was relatively complicated. If the SC concept could work successfully in that situation, it would likely be successful in a simpler setting. So I went ahead with the next phase of the pilot project.

B. Second Stage: First-Year Required Lawyering Course

This course is taught in four sections, each with 110 to 120 students. Each section uses the same teaching materials and the same final exam. I prepared a SC exercise for use only in one of the four sections. My principal goal was to assess the administrative utility of the SC concept when larger numbers of students were involved. I also wanted to assess the students' acceptance of—or resistance to—written feedback from actors.⁴⁹ Once again, the SC evaluations were not incorporated into the students' course grades.

I designed a SC exercise in which each student would interview a witness in the case in which they were immersed for the entire semester.⁵⁰ The students were working their way through the litigation file in a breach-of-contract case in which they were representing the party opposing a summary judgment motion. Among other things in the course, the students observed videotapes of the initial interview of their client and a deposition of a witness in the case. They also studied the applicable doctrinal law and developed legal and factual theories for their client. For the SC exercise, I prepared the profiles of two

48. Those initial results, though quite limited, are similar to those in a recent major medical school study comparing SP and medical faculty assessments. See Swartz et al, *supra* note 26. While the research did not establish the validity of the SP assessments for high-stakes exam purposes, it did establish "moderate" correlations between faculty and SP evaluations for interpersonal and communication skills. Jerry Colliver, a psychometrician, told me (in a telephone conversation, July 25, 2001) that a correlation coefficient of 0.2 is "small"; 0.5 is "midlevel"; and 0.8 is "large." For high-stakes use (e.g., a licensing exam), 0.8 is required; for low-stakes use, 0.5 is adequate.

49. Because of resource limitations, I was not able to compare faculty evaluations with those of the SCs, as I had done with the smaller first pilot.

50. I continued to use the term *standardized client* even though the exercise was a witness interview. It seemed to me then, and still does, a useful single term of art to describe the concept.

witnesses that should be interviewed by the lawyers in this case. They were persons peripheral to the controversy, but not unimportant in the development of a successful case theory. The SC exercise called on the student to perform interviewing skills and to synthesize the information obtained (or not obtained) with their developing case theories.

I hired fourteen actors to play the role of the two witnesses. Half the class interviewed one witness, and half the other. Working with the four teachers of the course, I prepared an evaluation checklist, which—like the SC exercise itself—was much simpler than the one used for the counseling exercise.⁵¹ It had three categories: information elicited, communication skills, and additional comments. Unlike the assessment form used in the first counseling exercise, the first part set forth a series of facts and sought only yes/no answers as to whether the student obtained the information. The second asked for a numerical rating on a 1-to-3 scale of a series of communication skills.⁵² Whereas the counseling exercise took forty-five minutes, we expected the witness interview to be completed within fifteen minutes (or at most twenty). In fact, the average duration was probably nine to twelve minutes. In part, the short interviews reflected the fact that many of these first-year students had never interviewed anybody for any purpose, and they lacked any sophisticated view of the task at hand.

One issue common to both the first- and the second-stage exercises was how to strike the correct balance between questions about factual details and assessments of communication skills. The goal was to ensure that the checklists did not become so detailed that they would reward excessive or obsessive thoroughness and devalue the more holistic aspects of interactive communication skills.⁵³ At the same time, there had to be enough questions so that the evaluation would be seen as comprehensive. In addition, as in the first stage, we asked the actors to complete an impressionistic “overall communication skills assessment” before they answered the detailed questions. Comparing the totals of the individual questions in the communication skills part with the overall assessments might indicate the degree of consistency in the SC evaluations. But, as with medical educators’ cautiousness about the use of the SP, the acceptance by the legal profession or the public of a layperson’s global

51. A copy of this witness-interview checklist is included as Appendix E at my Web site. *supra* note 10.

52. There also was a numerical overall assessment for the communication skills part. Following the SP empirical studies (see Cohen et al., *supra* note 13), I did not use the SCs’ overall assessments to establish the validity of the SC. The SC overall assessments would merely provide additional data that might be of some interest.

53. With respect to the SP, there is similar cautiousness about excessive reliance on the “objectification” of clinical skills. Cf. Devra S. Cohen et al., Psychometric Properties of a Standardized-Patient Checklist and Rating-Scale Form Used to Assess Interpersonal and Communication Skills, 71 Acad. Med. Supp. S87 (1996). (“Their concern was that these objective measures may focus on somewhat trivial and easily measured aspects of the clinical encounter, and that more subtle but critical factors in clinical performance may be overlooked or ignored.”)

evaluation of a lawyer's or law student's clinical skills may be pressing beyond realistic limits.

To enhance their learning experience, we asked the students to complete a checklist immediately after the exercise assessing their own performance.⁵⁴ The checklist was identical to the one the actors completed (except, of course, that it used the first person). The day after the exercise, we provided copies of the SC and student checklists to the students as well as to the four teachers.

While the experience was still fresh in their minds, the students were able to compare how they saw their performance with how their interviewee saw it. In the vast majority of the student interviews, there was tremendous similarity between the SC's evaluation and the student's self-evaluation.⁵⁵ This produced some powerful learning experiences for the students (as they told me in informal conversations). Further, the similarity of the two evaluations seemed to reinforce the students' self-critiques (positive and negative parts). The teacher, in turn, having also reviewed the pairs of checklists as well as videotapes of some of the interviews, was able to comment on some of the details of specific interviews in the followup class that discussed the exercises—which further reinforced the learning impact of the dual evaluations.⁵⁶

Finally, we distributed a form to the students asking them to assess the exercise anonymously.⁵⁷ The overall purpose of the assessment form was to gauge whether the students accepted the legitimacy of the exercise. We also wanted to know if they thought they were learning as a result of the experience generally and from the feedback of the actors specifically. The questions solicited student views on the management of the project as well as its educational value. The results were generally positive in every respect.⁵⁸ The students were very supportive of the continued use of this learning technique.⁵⁹

54. A copy of the student self-evaluation checklist is included as Appendix F at my Web site, *supra* note 10. See Susan C. Zonia, *Interns' Self-Evaluations Compared with Their Faculty's Evaluations*, 75 *Acad. Med.* 742 (2000).

55. This conclusion is based on my random scanning of the forms. Because of resource limitations, I was not able to complete a comparative statistical analysis of the two sets of evaluations.

56. Whereas all of the counseling exercises in the first stage of the pilot had to be taped (so they could be graded by the teachers independently of the SC experiment), here we taped only 10 to 15 of the 106 SC interviews. The taped examples were useful in our subjective assessments of the exercise and for possible selective showing in the classroom (which of course could only be done with extreme sensitivity). Once again, resource limits prevented us from completing a comparative statistical analysis of the SCs' and the teachers' evaluations.

57. This is in contrast to the evaluation checklists, which were coded so that a student's checklist could be identified and could be compared to the checklist of that student's SC. A copy of the student assessment form is included as Appendix G at my Web site, *supra* note 10.

58. My observations, generally, are based on both the second pilot and the third, which used the same SC exercises but involved 450 students rather than 110.

59. The student assessment form had three questions that addressed the issue of whether the use of SCs should be continued. Of the 103 students who completed a form for the second pilot project, between 76 and 89 percent of them gave either a 4 or a 5 (the highest scores), indicating a very positive response to the question of continued use.

They not only accepted, with a few exceptions, the validity of nonlawyer evaluations, but enthusiastically described the exercise (in supplemental written commentary) as one of the best parts of the course.⁶⁰ Again, I viewed many of the taped samples, observed the followup class discussion, and spoke informally with a number of the students and actors. My assessment of the SC exercise was equally enthusiastic. On the basis of those findings we went ahead with the third phase of the project—to expand the use of SCs to the full first-year class (450 students).

C. Third Stage: All Sections of First-Year Lawyering Course

Having thus established to my satisfaction the pedagogical value of SCs and the student acceptance of nonlawyer feedback, I wanted to assess the cost and the logistical and administrative hurdles if even larger numbers of students were involved. That was the goal of the third stage. The potential value of using SCs, essentially, is an economic one: can a valid teaching objective be accomplished that would otherwise be unattainable because of prohibitive costs? In this case, can the goal of providing high-quality individualized feedback on interpersonal lawyering skills be achieved through the use of SCs, and can it be done with large numbers of students and in a way at least approaching the quality and meaningfulness of personal feedback from a teacher?⁶¹ This third phase of the experiment tried to answer these questions.

We used the same witness-interviewing exercise that we had used for the single-section pilot project, making only minor changes in the role-play instructions and the evaluation checklists. We used 36 actors for 457 students. The students conducted the interviews during the regularly scheduled class periods over the course of four days, just as we had done for the single section. A few additional facts are worth noting. First, we used seven actors from the second pilot to reprise their roles. Second, we had two two-hour training sessions for the actors, as opposed to three sessions in the prior pilot. And all four of the teachers trained a group of (roughly) eight or nine actors in the two sessions; all of the actors attended both sessions except for the seven repeaters.

We again solicited student feedback on the SC interviews, and the responses continued to be quite positive.⁶²

60. It should be noted, however, that while the students were generally positive, there was a range in their responses. For example, in the third pilot their ratings for continuing the SC exercises were higher (question 6: 4.41 to 4.73) than their grading of SC evaluations (question 8: 3.76 to 4.15). These are among the many issues that call for further data collection.

61. Whether the teacher is a full-time faculty member or a paid adjunct, those costs would far exceed the cost of the SC.

62. All students were given the assessment forms in the class following the witness interviews and were asked to complete the form. It was not required. In the third pilot 253 of 456 students (57%) completed the assessment forms. But in one of the four sections the teacher had his students complete the form in class. He got an 84 percent return; the other three sections had returns of 63, 42, and 33 percent. The students' scoring differences among the four sections, however, were relatively insignificant for all nine of the questions on the assessment form. The scores also were quite similar to those for the single-class second pilot. With respect to the final pilot, the average score for each of the nine questions for the entire class

In addition to compiling the quantitative data summarized above, I undertook several other methods of subjective evaluation. Again I viewed a number of student interviews,⁶³ and I read all of the students' written comments on the exercise. I also reviewed many of the actors' evaluation checklists and compared them to the students' self-evaluation checklists. To observe how the interviews were integrated into the course, I again sat in on the classes that followed the students' interviewing. Finally, I spoke informally with perhaps a dozen students as well as the teachers in whose classes the interviews took place. On the basis of all these various means of evaluation, I reached the same generally positive conclusions that the students did: namely, that the students had learned a lot from a short exercise and enjoyed it immensely.

IV. The Potential Value of the Standardized Client in Legal Education

The goal of our experiment was to explore the potential educational value of SC exercises. Could the SC be used effectively and efficiently to provide constructive feedback to students on skills performances? Whether the SC could or should be used for high-stakes exams, or for an important graded exercise in a course, were distant concerns at most. The results of this three-stage pilot seem to me to strongly support the educational value of using the SC for the purpose of providing meaningful feedback to law students and perhaps even for the purpose of providing a pass/fail grade for a SC performance that is a small part of a course grade.⁶⁴ For these limited purposes, the issues that remain are ones only of cost and administration (which I discuss below).

The SC exercise enables the student to apply the law in context. It can hone analytical as well as interpersonal skills. Equally important, the SC can be used to enhance our teaching with the least imposition on, or displacement of,

is in the first column below; the comparable average scores for each separate section, as well as the single class from the second pilot, are in the columns to the right:

| Question number | Third pilot average (all classes) | Section 1 | Section 2 | Section 3 | Section 4 | Second pilot (single class average) |
|-----------------|-----------------------------------|-----------|-----------|-----------|-----------|-------------------------------------|
| 1 | 3.56 | 3.48 | 3.45 | 3.88 | 3.60 | 3.77 |
| 2 | 3.44 | 3.13 | 3.30 | 3.77 | 3.52 | 3.61 |
| 3 | 3.78 | 3.78 | 3.64 | 3.56 | 3.65 | 3.78 |
| 4 | 4.29 | 4.09 | 4.06 | 4.56 | 4.47 | 4.22 |
| 5 | 4.29 | 3.90 | 4.10 | 4.56 | 4.36 | 4.14 |
| 6 | 4.50 | 4.41 | 4.21 | 4.73 | 4.63 | 4.41 |
| 7 | 4.43 | 4.38 | 4.30 | 4.66 | 4.47 | 4.40 |
| 8 | 3.90 | 3.89 | 3.76 | 4.15 | 4.05 | 4.05 |
| 9 | 4.51 | 4.47 | 4.51 | 4.62 | 4.47 | 4.31 |

63. I observed roughly two dozen randomly selected student interviews, either live through closed-circuit TV or via videotape.

64. For any of these uses of the SC, even these low-stakes uses, it is desirable and probably necessary to compile additional empirical data that will either support or reject the very preliminary findings that are discussed in this essay. With respect to our plan at NYLS to further expand our use of the SC, we certainly intend to make every effort to design and implement a much more comprehensive data collection.

anything we are now doing.⁶⁵ The SC exercises are done outside the classroom, and could even operate outside the traditional course structure if necessary. What are the potential educational benefits of the SC?

A. Large-Class Discussions

How to engage students meaningfully in a large-class discussion of the law is a continuing challenge for most teachers. The first-year lawyering class described above provides a good illustration. I had previously observed the class being taught without using the SC exercise (and I also have taught the class). I then observed the same class taught (separately) by four of my colleagues after the students had completed the SC witness interview. The doctrinal purpose of the post-SC class was to examine the contribution, if any, that the witness could make to reinforcing (or undermining) the client's case theory. After conducting the SC interviews, the students struggled much more intensely with the pluses and minuses of the witness's potential value to our client's case. The collective discussion was much richer than it had been when the students had not had the benefit of the SC experience, when they had not been engaged so directly and intimately with the witness. The nuances of the legal and factual theory became much more real to them. They were able to connect the factual ambiguities and the personal, practical, and psychological issues arising out of the witness interview to their understanding of the applicable law in a much more sophisticated fashion. More specifically, they were able to move from the point of contemplating possible facts to the challenge of obtaining those facts, then analyzing concretely how the actual facts fit their legal theory, and finally determining how their legal theory might help or hurt their client. Because the students had taken the exercise seriously, they were genuinely engaged in the doctrinal discussion that tied the witness facts to the strengths and weaknesses of the legal case.

Similar results, it seems clear to me, are achievable in other traditional classroom situations, certainly not limited to first-year classes. For example, assume a trusts and estates class is assigned a leading appellate court opinion on the topic of the competency of an individual to change a will, together with a later trial court opinion on the same topic. The teacher then uses an exercise in which the SC is the client in the second case (i.e., the client who lost at the trial level contesting the validity of a will change). The student must counsel the SC on the pros and cons of appealing. To conduct the counseling session, the student must know how to explain the law and apply it to the specific circumstances of that client. Comparable results would seem to be achievable even in ethics courses, where simulation has proven to be quite effective.⁶⁶ Whether it is trusts and estates or ethics, the SC could take the more typical counseling simulation experience one or two steps further; for example, the SC could provide feedback on the organization of the client meeting or the

65. A frequent retort of traditional law teachers to the suggestion that simulation should be used in the large class is that it takes up valuable class time and limits coverage. The SC addresses that concern; its use need not take up any class time whatsoever.

66. See, e.g., Robert P. Burns, *Teaching the Basic Ethics Class Through Simulation: The Northwestern Program in Advocacy and Professionalism*, 58 *Law & Contemp. Probs.* 37 (1995).

clarity of the options presented. In turn, the post-SC class discussions of the meaning of the applicable case law will likely be richer, and the students will be much more engaged than if they had not counseled the SC.

B. An Additional Method of Teaching Clinical Skills

One-on-one teacher-student feedback and supervision with respect to the full range of lawyering skills is necessarily the most expensive form of education. For many reasons, it is also critical that such clinical teaching continue to be an integral part of legal education.⁶⁷ The SC offers an opportunity to supplement and expand what is now being done to prepare law students to be competent interviewers and counselors. In particular, the SC can be used in situations where lawyering skills are not now being taught. The trusts and estates exercise just discussed is a good example. While the SC exercise clearly enriches the doctrinal learning, it is equally clear that it can provide experiential learning in the skills of interviewing and counseling. SC evaluation is not a substitute for a teacher's feedback. But it clearly enhances the experiential value of the interview. The same could be said for most first-year classes. Most large classes are taught by teachers whose interests lie elsewhere than clinical/skills education.⁶⁸ The SC, if administered properly, is an ideal teaching tool for all categories of traditional large classroom courses.

Even without a word being spoken in the large classroom about any possible skills learning, the students would be ahead of those without the benefit of the SC. The students would be receiving written feedback after each exercise. They also could compare a self-evaluation with that of the SC. The traditional doctrinal teacher would only have to suggest scenarios (either from opinions assigned or from hypotheticals used in class) that could be transformed into SC exercises and then schedule the exercises in conjunction with the relevant topics.⁶⁹ If a teacher wanted to bring aspects of the skills or related ethical issues into the class discussion, all the better. But it would not be necessary.

SC exercises might also be used to supplement simulations in skills courses and even in clinics. Again—let me be explicit—the objective here would not be to reduce the amount of one-on-one feedback now being offered in clinical

67. A discussion of the justifications for clinical education is beyond the scope of this article. See, e.g., Section of Legal Education and Admissions to the Bar, American Bar Association, Report of the Task Force on Law Schools and the Profession: Narrowing the Gap, Legal Education and Professional Development—An Educational Continuum (Chicago, 1992) (known as the MacCrate Report). Briefly, live-client clinical courses enable teachers to remain connected to the real world and the skills demands of interacting effectively with clients, lawyers, and the courts; provide a laboratory to examine alternative theories for improving skills-teaching methodologies; offer a way for students to learn firsthand of the inequities in our legal system; provide a way in which students can most powerfully learn the ethics challenges they will confront as lawyers; and enable the law school to make a pro bono contribution, in the form of legal services, to the pursuit of justice in our legal system.

68. This is true even of the many faculty who support clinical/skills education—as long as it done by others.

69. This assumes, of course, an adequate administrative structure (discussed below).

courses in most schools (indeed, it is inadequate in many institutions), but rather to supplement the direct supervision now being given. For example, if a civil rights clinic regularly gives a student a chance to do one simulated interview (followed by individual faculty feedback) before seeing a real client, giving that same student a second SC interview before the teacher's feedback session would enrich the faculty-student discussion and the critique.

C. A Vehicle for Assessing the Value of Student Self-Critique

Self-assessment has been a central feature in clinical education. The SC offers a way to assess how accurate or useful student self-critiques are. The pilot project discussed above demonstrates that the SC could facilitate the collection of a quantity of data on this important learning tool.⁷⁰

D. A Means to Improve Our Testing

The SC pilot project I describe in this essay was conducted in the context of what started out as an experimental course.⁷¹ While learning doctrinal law was an important part of the lawyering course, its primary objective continues to be to introduce students to what lawyers actually do. In particular, we try in this course to teach students how lawyers get facts and what they can and should do with them once they get them. The course's introduction to interviewing, fact investigation, and counseling is intended to accomplish that goal. The use of the SC has been ideal in helping to accomplish that purpose.

As part of its experimental nature, the course uses a performance test with a video component for its final exam. Building on the litigation file that the students have worked with throughout the semester, we give the students additional file materials (memos, deposition transcripts, etc.) and new case law, and we ask them to synthesize the new materials into the semester-long file and write an analysis. We also ask them to view a tape of a lawyer typically counseling the client on what the new law and facts mean for the client, and then to write a critique of the lawyer's performance, both as to the law and as to the skill with which it is applied. Again, the students' experience with the SC exercise seems to have enhanced their ability to produce good skills critiques. This is still another area ripe for further empirical study.

V. The Future for the Standardized Client

Two key aspects of future SC use are integrally related. On the one hand, the SC requires an unusual administrative structure to be effective and economically viable. That structure calls for more exercises to satisfy the economy

70. Medical researchers, for example, knowing that it is important for physicians to acknowledge when they do not know something, examined medical students' self-assessments and found no discernible difference between cognitive self-assessments and performance-based self-assessments. They concluded that since "self-assessment is more analagous to a personality characteristic than it is to problem-solving behavior," it is not clear whether self-assessment skills can be taught. Fitzgerald et al., *supra* note 11, at 740.

71. The lawyering course has now been taught seven times, and the designers of the course consider it a success, in part due to the SC exercises.

of scale. The larger number of SC exercises fits neatly into the need for empirical data to support the validity of the SC and, more generally, the propriety of layperson feedback and evaluation of law students. The mutual interdependence of these two factors has been amply demonstrated in the Mt. Sinai Medical School Morchand Center in New York City. It is the model I propose for legal educators.

The Morchand Center is supported by a consortium of the eight medical schools in the metropolitan area. It has a staff headed by a medical director and an educational director.⁷² Over the course of the year, roughly 1,500 medical students from all eight schools conduct simulated standardized patient exercises at the center. Large numbers of actors are hired, and a substantial number of them play their roles again and again. Those repeat players no longer require training time, and the costs go down further. The center has seven interview/examination rooms, each one equipped with an unobtrusive video camera, each connected to a control room with seating for fifteen. A teacher (with or without students present) can view one or more of the sessions live or watch a tape well after the time of the interview. Each medical school makes an independent decision about how to use the SP experiences. Some require the students to receive a passing mark in order to graduate; others simply assign the exercises for the benefit of the student's learning.

The Mt. Sinai model certainly is ideally suited for a metropolitan area that has several law schools. But even in areas with only one or two law schools the model could work. An obvious initial obstacle is obtaining startup money to establish a physical facility and hire the staff minimally necessary to administer it. There also has to be at least one law faculty member who can play a central role. Once such a center is in place and operating, the Mt. Sinai example suggests that it can be maintained at a relatively low cost per school if all the participating schools share in the costs. And it is possible that other professions might even be able to use such a facility and share in the expenses.

Not only is a large-scale operation optimal to achieve the maximum economy of scale,⁷³ but it is also crucial for ease of access by traditional classroom teachers. If a torts teacher and a contracts teacher want to assign an exercise in the same week, or even on the same day, they do not want to hear about scheduling conflicts. The concept can probably work in the way that I posit only if there is very easy accessibility by teachers who otherwise would not use the SC device.

Actors, I have learned, have been available for SP exercises all over the country, and not just in New York City or Los Angeles. They could be available for SC exercises as well. An efficient center could even be established by a

72. While the Morchand Center initially was the beneficiary of a large grant from the Macy Foundation, it is now financially supported by each of the participating medical schools. That crucial initial grant enabled Mt. Sinai to construct the physical plant for the center.

73. The SC can also work on a much more limited scale, just not as efficiently. Indeed, now that we have used the SC several times at NYLS for all 450 first-year students, it is administered relatively easily.

single school; it simply is a more costly initial investment and would entail greater maintenance expenses. Single medical schools have done it.⁷⁴

But cooperative ventures among law schools are rare. Aside from library-sharing arrangements, none come to mind. In a locale like New York City, there is no shortage of law schools that might benefit from such a cooperative effort. I began preliminary inquiries in my area, and got very positive initial responses. There are several key questions. How much money is needed to establish such a center? How does the amount compare with what the alternatives would cost? What are the quality differences among the various ways in which to provide feedback to students on role-playing exercises?

Without getting into the dollar specifics that would be required for such a center, certain propositions seem clear.⁷⁵ The costs of providing individual feedback to students from full-time faculty would be prohibitively expensive. The costs of providing feedback by trained and paid adjunct faculty would be less, but considerably more expensive than the SC.⁷⁶ In my experience with volunteer lawyers in the moot court context and with student role-players (two possible substitutes), the quality of the feedback provided by such alternatives to SCs would be inferior to that given by SCs.⁷⁷

Another issue is whether there are paper-and-pencil tests that might be an adequate substitute for the more expensive SC evaluation. For example, a law student might be presented with a client file, then asked to view a videotaped lawyer-client counseling session, and then asked to complete a written critique and analysis of the lawyer-client interaction. I call this a video performance test.⁷⁸ At best, however, such a written test is only a supplement to or a partial substitute for an evaluation of an actual performance by the test taker. All of this assumes, of course, that the SC evaluation would be sufficiently valid to be included as part of a course grade—a large assumption. This is still another

74. The University of Arkansas has a facility with 10 examination rooms and a database of 200 SPs; 100 of them work regularly (more than once a year). Students in pharmacy and nursing, as well as medicine, participate. Eric Harrison, *Medical Practice Actors Become Standardized Patients to Help Health Care Students, Professionals Polish Their Skills*, *Ark. Democrat-Gazette*, April 10, 2000, at E1. At the University of Kentucky Medical School, the SP program has 100 SPs ranging in age from 7 to 83. Barbara Isaacs, *Actors Play Sick for Doctors*, *Chi. Tribune*, July 11, 1999, at 8.

75. I have not seen the budget for the Morchand Center. I know that the cost of producing 450 SC exercises in our first-year class is about \$17,500. But it is not possible to extrapolate from this figure what the costs of a center would be.

76. While the SC is less expensive, it is not cheap. Ironically, one of the reasons for the use of SPs is that too many medical students were completing medical school with little or no feedback from clinical faculty. The thought was that SP feedback was better than none at all. Or, putting it slightly differently, it would be less expensive than providing one-on-one faculty feedback. One supporter of SPs notes that it would be the wrong message to suggest that SPs should be used to replace such teacher-student interactions because they are cheaper. Stillman, *supra* note 14, at 468.

77. My experience with student role-players is mixed at best. While they are less expensive than actors, they are less reliable and simply not as good at playing a role. Some medical schools recommend using students as SPs to save money. See, e.g., V. Andres Sassoon et al., "Teach 1, Do 1 . . . Better": Superior Communication Skills in Senior Medical Students Serving as Standardized Patient-Examiners for Their Junior Peers, 74 *Acad. Med.* 932 (1999).

78. See Grosberg, *supra* note 33. Is there an analog in medical education? I have not seen any references to such an alternative to either the SP or physician-observed examinations.

potential research area to explore. It would entail some kind of comparison between the lawyering performances of students who perform SC exercises and those of students who only complete written tests.

The ultimate question is whether enough schools would conclude that the benefits of the SC to their students justify their portion of the initial as well as the continuing shared costs to establish and maintain such a center. Here, again, the political value of empirical research (assuming it is supportive of SC use) is that it will strengthen the argument for law school cooperation to establish a SC center.

The data collection objectives are certainly achievable in the context of a law school model based on the Mt. Sinai Center. Assuming the financial support for such empirical research, this model at least presents the vehicle for beginning to collect the kind of data that medical researchers have been collecting for more than twenty years.⁷⁹

VI. The Standardized Client and High-Stakes Testing

The most problematic use of the SC is on bar exams. I raise it briefly as a topic for discussion only because the medical analog already is part of the licensing exam, at least for foreign-educated applicants. Also, the SC has proved useful in reinforcing the effectiveness of a video performance test in our first-year lawyering course. To the extent such a video performance test or even a SC interview might ever be a part of the bar exam, the expanded current use of SCs might indirectly affect the course of development of bar exams. Realistically, however, the inclusion of a SC on a bar exam, or even as part of a law school course grade, is light years behind medical schools and the medical licensing exams.

Before the SC can or should be even considered as a possible component of a bar exam, much empirical study must be done. The kind of work that our medical colleagues have been doing for twenty years has not been begun by legal educators. If a center were established as I have suggested, that research could start.

Nevertheless, the logic of including the SC as part of the process of being admitted to practice law seems theoretically unassailable. It is a way of testing for competence in interviewing and counseling, two skills critical to competent lawyering. This is the same rationale that has impelled the medical licensors to include the SP on licensing exams. If a bar examiner cannot watch and evaluate a bar applicant conducting an interview with a real client, or even a simulated client, the SC could be an appropriate alternative. That, indeed, is exactly what the medical examiners have tentatively concluded with respect to the skill of interviewing and examining a patient. Our bar examiners now do almost no evaluation of comparable lawyering skills.⁸⁰ Yet we purport to certify

79. While it is much easier with a multischool center, even a single law school can begin empirical research. At NYLS we are expanding our SC efforts, and we hope to get the funding to conduct meaningful SC empirical research.

80. The adoption of the performance test as a small part of bar exams is a step in the direction of expanding the array of skills tested on a written bar exam; see Grosberg, *supra* note 33.

to the public that persons who pass the bar exam are prepared and competent to practice law. There certainly is room for empirical study in this area.

* * * * *

There is little if any disagreement among legal educators that giving students the opportunity to apply the law in a simulated context is a useful learning experience. There also is general agreement that students benefit from individual feedback on their performance in such a simulation. It is a fact, however, that not a lot of such feedback is provided to most law students. The standardized client concept offers a way to address that shortcoming in our standard curriculum, and to do so at a relatively low cost. The SC could enrich traditional doctrinal teaching as well as provide experiential skills learning. Following the lead of our medical school colleagues, we ought to move to implement the idea, at least on an experimental level and for the purpose of collecting empirical data. Whether we ever will, or should, consider the technique for course grading or for such high-stakes tests as the bar exam are questions that need not be answered at this point.