

Effect of Communications Training on Medical Student Performance

Michael J. Yedidia, PhD

Colleen C. Gillespie, PhD

Elizabeth Kachur, PhD

Mark D. Schwartz, MD

Judith Ockene, PhD

Amy E. Chepaitis, MBA

Clint W. Snyder, PhD

Aaron Lazare, MD

Mack Lipkin, Jr, MD

DESPITE WIDESPREAD acknowledgment of the importance of improved patient-physician communication,^{1,2} teaching of communications skills has not been systematically integrated into most medical school curricula^{3,4} and has not been subjected to evaluation across different schools.⁵ Furthermore, few standardized assessment instruments specifically measure students' communications performance.⁶⁻⁸

METHODS

Intervention

Three medical schools—New York University, University of Massachusetts, and Case Western Reserve University—initiated a curriculum in 2000-2001 to teach a common set of communications competencies during the third clerkship year, based on an established educational model. Although implementation at the 3 schools reflected differences in their culture, history, and curricular organization, the interventions shared 4 common characteristics. First, they used a documented model for teaching communications skills^{9,10} that relies on experiential teaching modes;

For editorial comment see p 1210.

Context Although physicians' communication skills have been found to be related to clinical outcomes and patient satisfaction, teaching of communication skills has not been fully integrated into many medical school curricula or adequately evaluated with large-scale controlled trials.

Objective To determine whether communications training for medical students improves specific competencies known to affect outcomes of care.

Design and Setting A communications curriculum instituted in 2000-2001 at 3 US medical schools was evaluated with objective structured clinical examinations (OSCEs). The same OSCEs were administered to a comparison cohort of students in the year before the intervention.

Participants One hundred thirty-eight randomly selected medical students (38% of eligible students) in the comparison cohort, tested at the beginning and end of their third year (1999-2000), and 155 students in the intervention cohort (42% of eligible students), tested at the beginning and end of their third year (2000-2001).

Intervention Comprehensive communications curricula were developed at each school using an established educational model for teaching and practicing core communication skills and engaging students in self-reflection on their performance. Communications teaching was integrated with clinical material during the third year, required clerkships, and was supported by formal faculty development.

Main Outcome Measures Standardized patients assessed student performance in OSCEs on 21 skills related to 5 key patient care tasks: relationship development and maintenance, patient assessment, education and counseling, negotiation and shared decision making, and organization and time management. Scores were calculated as percentage of maximum possible performance.

Results Adjusting for baseline differences, students exposed to the intervention significantly outperformed those in the comparison cohort on the overall OSCE (65.4% vs 60.4%; 5.1% difference; 95% confidence interval [CI], 3.9%-6.3%; $P < .001$), relationship development and maintenance (5.3% difference; 95% CI, 3.8%-6.7%; $P < .001$), organization and time management (1.8% difference; 95% CI, 1.0%-2.7%; $P < .001$), and subsets of cases addressing patient assessment (6.7% difference; 95% CI, 5.9%-7.8%; $P < .001$) and negotiation and shared decision making (5.7% difference; 95% CI, 4.5%-6.9%; $P < .001$). Similar effects were found at each of the 3 schools, though they differed in magnitude.

Conclusions Communications curricula using an established educational model significantly improved third-year students' overall communications competence as well as their skills in relationship building, organization and time management, patient assessment, and negotiation and shared decision making—tasks that are important to positive patient outcomes. Improvements were observed at each of the 3 schools despite adaptation of the intervention to the local curriculum and culture.

JAMA. 2003;290:1157-1165

www.jama.com

Author Affiliations: Center for Health and Public Service Research, Robert F. Wagner Graduate School of Public Service (Drs Yedidia and Gillespie and Ms Chepaitis), and School of Medicine (Drs Schwartz and Lipkin), New York University, and Medical Education Development (Dr Kachur), New York City; University of Massachusetts Medical School, Worcester (Drs

Ockene and Lazare); and Case Western Reserve University School of Medicine, Cleveland, Ohio (Dr Snyder).
Corresponding Author and Reprints: Michael J. Yedidia, PhD, Center for Health and Public Service Research, Wagner Graduate School of Public Service, 726 Broadway, Fifth Floor, New York, NY 10003 (e-mail: michael.yedidia@nyu.edu).

simultaneous attention to knowledge, skills, and attitudes; and learner-centered educational approaches. Second, the interventions addressed a common set of competencies grounded in relevant literature¹¹ and organized around the structure, sequence, and function of the medical interview.¹² Third, each school dedicated time to teach communication skills, integrated with the clinical material of third-year required clerkships. Finally, each school supported the intervention with formal faculty development to ensure effective guidance and individualized feedback to students.^{13,14}

Although each school was free to develop its own specific content areas, the curriculum at each of the schools focused on promoting the same set of underlying competencies. These core skills included determining the reasons for the patient's visit, eliciting and understanding the patient's perspective, sharing information and providing education, negotiating and agreeing on a plan, and achieving closure. The teaching approach was also standardized across the curricula.^{9,10} It relied on an iterative pro-

cess that incorporated demonstration of interviewing skills by clerkship directors and faculty, experiential learning techniques (eg, student interaction with standardized patients [SPs]), individualized feedback, and student self-reflection on how their attitudes and values affected their performance. The core skills were taught as integral to mastery of designated clerkship topics at each school (TABLE 1).

The intervention was implemented in 2000-2001, and the entire third-year class at each school participated (n=373 for all 3 schools). Before the intervention, communications training had been confined mainly to the pre-clinical years. Attention to communications teaching in the clerkships at 2 of the schools had been episodic, coinciding with bedside teaching, while at the third school it had been more prominent but was not coordinated to introduce and reinforce an integrated set of skills.

Evaluation Design

The intervention cohort (class of 2002) was the first cohort that was exposed

to the new curricula. A random sample of all students at each school from the classes of 2001 and 2002 comprised the comparison and intervention cohorts, respectively. At the largest school, one third of all students were selected while at the smallest school, 82% were sampled. For both cohorts, each student was assessed using identical instrumentation prior to the beginning and at the end of the third year. Both cohorts at each school received the same preclinical curriculum.

Measures

Both the pretest and posttest assessment consisted of an identical 10-station objective structured clinical examination (OSCE) that assessed communications skills known to affect patient outcomes.¹⁵ To develop the OSCE, we defined 5 patient care tasks grounded in the literature on patient-physician communication. For each task, we further defined specific communication skills that have been associated with improved patient outcomes. We then developed specific checklist items for SPs to use in assess-

Table 1. Designated Topics for Teaching Core Communication Skills Within Clerkships at Each School

Clerkships	Topics		
	School A	School B	School C
Medicine	Counseling Sexual health Ethics	Chronic pain Addiction	Alcoholism Cultural issues End-of-life care
Surgery	Oral presentation Ethics	Patient education	Bad news Informed consent Patient education
Family medicine/ambulatory care	Medical informatics Sexuality Ethics	Chronic pain Addiction	Behavior change
Obstetrics and gynecology	Reproductive choice Sexual history Lesbian health Bad news Genetic counseling Ethics	Domestic violence	Sexuality Family violence Genetic counseling
Pediatrics	Bad news Genetic counseling Sexuality/sexual orientation Nutrition Ethics	Ethics Cross-cultural issues	Asthma Adolescent history
Neurology	None	Bad news	Dementia Headaches
Psychiatry	Sexual dysfunction Ethics	Angry patients	Hostile, seductive, or depressed patients
Special sessions (eg, interclerkships)	Nutrition Cultural competence End-of-life care	Cross-cultural issues	None

ing student performance. These 21 operational definitions (in addition to the SPs' global assessments of the students' competence) comprised our major dependent variables (TABLE 2). These specific skills have been associated with such outcomes as increased patient satisfaction,^{16,20,23,26,31-33} decreased worry,^{21,31} improved retention of information,^{22,23} more comprehensive medical histories,^{17-19,29,30} improved patient adherence to treatment plans,²⁰ reduction in symptoms,^{24,25} improved physiological outcomes,^{24,25,27,28} higher functional status,^{24,28} and reduction in malpractice claims.³⁴ Finally, we developed 10 clinical cases affording repeated assessments of these patient care tasks.

Local evaluators from each school collaborated with the central evaluation team to develop the 10 cases as well as the checklist items. Faculty members responsible for curriculum design and faculty development at each school, however, were unaware of the specific elements of the evaluation until after completion of the study. This helped ensure that the curricula would be guided by commitment to a broad spectrum of competencies and regard for local priorities, rather than a desire to "teach to the test."

Development of the cases was based on the priorities elicited from faculty at each of the 3 schools. TABLE 3 outlines the 10 cases (stations) comprising the OSCE and the competencies to be assessed with regard to each. Two of the tasks (relationship development and maintenance and organization and time management) are generic and were assessed in all 10 cases. A distinct set of cases was designed to assess the other 3 patient care tasks (ie, assessment of the patient's problem, patient education and counseling, and negotiation and shared decision making). At the completion of the OSCE, students were asked to report on whether the cases were realistic and representative of their actual clinical behaviors.

At each of the 10 OSCE stations, students were given 10 minutes to interact with the SP, then SPs were given 5 minutes to complete their evaluation for

a total of 2.5 hours of assessment per student. During the 5-minute intervals, students reviewed fact sheets and instructions for the next station that summarized relevant clinical knowledge for the task to be undertaken and outlined the objectives to be accomplished. The fact sheets were written to minimize any effect on communications performance of differences in clinical knowledge among students. For example, in 1 case, the patient was hospitalized for a bleeding ulcer for which heavy alcohol use was suspected as an aggravating factor. The student was instructed to take on the role of a first-year resident in medicine, talk to the patient about his alcohol problem, and explore alcohol treatment options with

him. The instructions provided socio-demographic characteristics of the patient, prior information elicited from the patient and his wife regarding his alcohol use, and his responses to the CAGE questionnaire, a standard alcoholism screening tool.³⁵ The fact sheet included diagnostic criteria for alcohol problems; information on the potential relationships among alcohol use, gastrointestinal bleeding, and ulcers; the value of hospitalization as leverage for motivating a patient to confront the need for lifestyle changes; and common strategies for counseling a patient about an alcohol problem.

Each school recruited and trained its own SPs, using common protocols and standard-setting videotapes for case por-

Table 2. OSCE Dimensions: Patient Care Tasks, Communication Skills, and Checklist Items

Communication Skills and Evidence of Impact on Patient Care	SP Assessment Checklist Items
Assessment of Patient's Problem and Situation	
Uses appropriate questioning techniques ^{16,17}	Asked 1 question at a time Moved from open- to closed-ended questions
Seeks understanding of patient's beliefs ¹⁸	Asked for SP's views about problem/illness
Checks accuracy of understanding ¹⁹	Summarized SP's story to check its accuracy
Patient Education and Counseling	
Elicits patient's perspectives ^{20,21}	Asked for SP's views of need for action, benefits, and barriers
Provides clear instructions and explanations ²²	Provided information in segments that were small enough to be absorbed Repeated or summarized important information
Assesses patient's understanding ²³	Ascertained what SP understood
Negotiation and Shared Decision Making	
Surveys problem and delineates options ^{24,25}	Summarized the problem or condition Presented full range of options
Arrives at mutually acceptable solution ^{24,26-28}	Invited SP's ideas and feelings about the treatment plan Made statements of partnership
Relationship Development and Maintenance	
Permits patient expression ^{29,30}	Allowed SP to speak without interruption Acknowledged SP's emotions
Communicates supportive attitude ^{31,32}	Was not judgmental Communicated concern and willingness to help
Uses appropriate style ³³	Explained unfamiliar jargon Nonverbal behavior enhanced communication
Organization and Time Management	
Organizes the encounter and concludes it effectively ³⁴	Set the agenda with SP input Moved through the encounter in an efficient manner Brought the encounter to some conclusion (eg, summarized session, elicited final questions)

Abbreviations: OSCE, objective structured clinical examination; SP, standardized patient.

trayal and assessment. Detailed written instructions were given to the SPs describing the clinical aspects of their case and specific directives for portraying it. For example, for the case on problem drinking, these instructions included a detailed biography relating to stresses at work and at home, the context and extent of alcohol use, and relevant aspects of family history; examples of defensive and aggressive responses to students' references to problems with alcohol; specific non-verbal cues to communicate to the student (indicating denial and hostility); and expectations for the visit (hospital discharge) and an opening statement ("Hi Doc; so I'm going home soon?"). A central training coordinator provided feedback on each SP's performance to enhance uniformity in presentation of the OSCE across the 3 schools.

We calculated OSCE scores as the percentage of items for which SPs gave the students full credit. In addition, SPs assigned students a global rating indicating whether the SP would recommend the student to a friend who was seeking a clinician with excellent communications skills (4-point scale). There is evidence that global ratings may be more valid than checklist scores in some circumstances.³⁶⁻³⁸ Seven mea-

asures are addressed in the analysis: the overall OSCE score; the scores for those items assessing development and maintenance of the relationship and organization and time management of the visit; the average score for each subset of cases addressing patient education, patient assessment, and negotiation and shared decision making; and the global recommendation. Reliability (Cronbach α) of these measures in pretest and posttest administrations of the OSCE across the 3 schools ranged from 0.50 to 0.75, with the average for the overall reliability for the OSCE score exceeding 0.70.

We were concerned that variations in SPs' portrayal of cases as well as anomalies in their ratings of student performance could diminish the reliability of students' OSCE scores, particularly when more than 1 SP was used for the same case. Such variations are often assumed to be randomly distributed throughout the OSCE and therefore may balance out when computing overall OSCE scores.³⁹ However, our emphasis on comparisons between cohorts over time at 3 sites necessitated special attention to ensuring that these threats to reliability did not confound the findings.⁴⁰ To establish the consistency of SP case portrayal over time, we developed measures of the level of emo-

tional intensity and the accuracy of the content of each case and applied them to videotaped performances from the first OSCE. We found that individual SPs were consistently accurate over time and that they were also consistently intense in their portrayals. However, there were variations in emotional intensity among different SPs portraying the same case. Therefore, we created profiles of portrayals for each case at the time that the OSCE was first performed at each school. These profiles were used in recruiting and training SPs for subsequent OSCEs to replicate the mix, thereby standardizing the level of emotion to which students were exposed across repeated assessments.

Given that SPs were recruited and trained at 3 different schools during 4 cycles over a 2-year period, we sought to set a standard for ensuring uniformity of assessment. To assess intrarater reliability for a subset of encounters, we compared SPs' ratings of students during live interactions with their subsequent ratings (at least several months later) of a videotape of that encounter. We found a high correlation between these 2 sets of scores ($r=0.83$; $P=.001$), indicating that individual SPs were consistent in how they rated students' performance over time.

Table 3. Objective Structured Clinical Examination Cases

Case Type*	Cases	
	Clinical Problem	Description
Assessment of patient's problem and situation	Impotence	Young man with psychogenic impotence
	Fever phone call	Young mother of child with fever calls late at night to find out what to do
	Fatigue	Elderly man recently lost spouse; very depressed affect
Patient education and counseling	Alcohol problem	Man hospitalized for bleeding in upper gastrointestinal tract meets criteria for alcoholism
	High-risk smoker	Long-term male smoker wants to quit
	PPD positive	Woman with negative radiograph but positive PPD test result
Negotiation and shared decision making	Medical mistake	Patient experiencing dizziness finds out she was given the wrong medication
	Herniated disk	Elderly patient who requires magnetic resonance imaging requests hospitalization for procedure
	Informed consent	Mother of adolescent with broken leg refuses blood transfusion for religious reasons
	Bad news	Uninsured patient with few financial resources receives follow-up to abnormal mammogram result

Abbreviation: PPD, purified protein derivative of tuberculin.

*Relationship development and maintenance and organization and time management were assessed in all cases.

To determine whether differences in rating styles among SPs were likely to confound our findings, we established a benchmark rating for use as a standard of comparison for identifying which SPs were particularly lenient or stringent in their ratings. Two of the authors (C.C.G. and A.E.C.) used explicit behavioral criteria for assessing each of the 21 checklist items as well as the global recommendation and independently applied them to 167 representative videotaped encounters. Their ratings of this set of cases were highly correlated ($r=0.85$) and did not vary systematically. The benchmark was then computed as the average of their ratings of each encounter. To establish the validity of these benchmarks, they were compared with those of a panel of 9 experts from the 3 schools, who independently assessed 92 of these same encounters ($r=0.72$).

The benchmark ratings were then compared with each SP's actual rating of the encounter to determine whether that SP deviated substantially from the standard. An SP rating was considered easy or difficult if it differed from the benchmark for a given encounter by more than 15%, a threshold that was chosen as unlikely to be due to chance across the multiple samples. The remaining SPs were considered accurate. Applying these criteria to the population of SPs across the 3 schools, we established that SPs were generally consistent over time in their rating styles, in accord with the analysis of intrarater reliability. However, there was considerable variation among SPs, and the mix of difficult, easy, and accurate SPs who were used for the OSCEs over time disproportionately favored easier raters. To adjust for this bias, we used a correction factor. For each of the measures, the score assigned to a student by an SP classified as difficult was increased by the mean difference between the scores assigned by all difficult raters and the scores assigned by accurate raters. Similarly, the difference between the mean of accurate and easy raters was subtracted from the score of an easy rater. For most students, this correc-

tion resulted in only minor modifications of their raw scores; on average, across all OSCEs, scores were adjusted by less than 2% with the largest correction made to the comparison cohort's pretest scores (12% [SD, 9.0%]), the most difficult of the OSCEs. A variable that did not require adjustment, SP ratings of whether students achieved the objectives of each case, yielded findings that were similar to those associated with the corrected measures, conferring further confidence in our process of adjustment.

Analysis

We used analysis of covariance to adjust each posttest score by differences at pretest and a series of analyses of variance to test for the main and interaction effects of cohort (comparison or intervention) and school on the OSCE outcome variables. To minimize the possibility of a type I error associated with testing effects on multiple dependent variables, a multivariate analysis of variance was conducted first to test for the significance of the main and interaction effects on the combined outcome variables. Follow-up analyses of variance were conducted separately for each of our individual outcome measures, using a Bonferroni correction to maintain an overall $P<.05$ for the set of 7 analyses. The dependent variables consisted of the adjusted posttest scores for each of our 7 measures. Because of their interpretative appeal, analyses were also conducted using change scores (ie, subtracting the pretest from the posttest scores and comparing the change in performance of the intervention and comparison cohorts). Analysis of covariance was chosen as the main analytic approach because it is more powerful⁴¹ and controls more effectively for baseline differences in pretest scores and in demographic characteristics of the cohorts.⁴¹⁻⁴³ Use of both analytic approaches is intended to strengthen causal inferences as to the impact of the interventions given our controlled but nonrandomized experimental design. SPSS version 11.0 statistical software (SPSS Inc, Chicago, Ill) was used for all analyses.

RESULTS

On average, 78% of pretest students completed the posttest assessment across the 6 samples (ie, intervention and comparison cohorts at 3 schools), with a range of 66% to 89%; final sample sizes for the comparison and intervention cohorts across the 3 schools were 138 (38%) and 155 (42%), respectively. Possible response bias was explored by comparing pretest performance of students who did and did not complete the posttest; there were no significant differences. Differences among cohorts unrelated to the intervention were assessed by comparing key demographic characteristics (ie, sex, age, race/ethnicity, undergraduate major, undergraduate grade point average, Medical College Admission Test scores) of the intervention and comparison cohorts at each school. There was only 1 significant difference. At school A, students in the 2 cohorts differed by age ($P<.001$) and race/ethnicity ($P=.02$), variables that could influence communication performance. Consequently, analyses of performance involving school A included adjustments to control for these demographic differences in cohorts. Responses from student questionnaires administered immediately following completion of the OSCE provided evidence that the students viewed their interaction with the SPs as a valid indicator of their communications abilities: 99% of all participants across schools reported that the OSCE stations were very or somewhat realistic, and 89% reported that their performance interacting with SPs was similar to how they behave with real patients.

Although the communications skills of students in each cohort improved during the third year, students who were exposed to the intervention improved more than those who were not (FIGURE). The impact of the intervention on the combined outcome measures was significant ($F=88.79$; $P<.001$), and this impact differed by school ($F=27.36$; $P<.001$). Subsequent analyses of variance revealed that 6 of the 7 outcome measures were re-

sponsible for the significance of the main effect. Controlling for individual pretest differences, students in the intervention cohort significantly outperformed those in the comparison cohort on our overall OSCE score across all 3 schools ($F=72.00; P<.001$). Examination of the posttest means for the full sample (TABLE 4) shows that the intervention cohort's scores on the 21 communications tasks, adjusted by their baseline performance, averaged 65.4% correct, compared with 60.4% for their counterparts in the comparison

cohort (mean difference, 5.1%; 95% confidence interval [CI], 3.9%-6.3%; $P<.001$). In addition, exposure to the intervention was associated with a higher global recommendation ($P<.001$) as well as superior performance with respect to relationship development and maintenance (mean difference, 5.3%; 95% CI, 3.8%-6.7%; $P<.001$), organization and time management of the visit (mean difference, 1.8%; 95% CI, 1.0%-2.7%; $P<.001$), assessment of the patient's problem and situation (mean difference, 6.7%; 95% CI, 5.9%-7.8%; $P<.001$), and negotiating and sharing decision making with patients (mean difference, 5.7%; 95% CI, 4.5%-6.9%; $P<.001$). The magnitude of these effects can best be assessed statistically,⁴⁴ given that empirical determinations of meaningful effect sizes in communications assessment do not yet exist.⁴⁵ Our findings demonstrate effect sizes that range from moderate (eg, Cohen $d=0.45$ for global recommendation, suggesting that membership in the intervention cohort accounts for 6% of the variance) to large (eg, Cohen $d=0.90$ for overall OSCE score, suggesting that membership in the intervention cohort accounts for 20% of the variance in performance score).

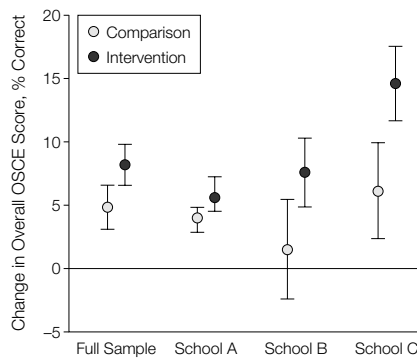
For 5 of the 7 outcomes, the interaction between cohort and site was significant (TABLE 5), accounting for the corresponding finding in the multivariate analysis of variance and suggesting that, for these outcomes, the impact of the intervention depended on the school in which the student was enrolled. Comparisons of adjusted posttest scores within school for each outcome shows that students in the intervention cohort consistently outperformed the comparison cohort, although the magnitude of these differences varied between schools. The mean differences in overall OSCE scores for each school ranged from 3.1% to 7.3%. For the subscores for which reliability at each school was sufficient to merit analysis, students in the intervention cohort outperformed their counterparts in the comparison cohort in every case; in 8 of the 13 instances, the difference was statistically significant.

Differences in change scores are presented in the Figure. Across all 3 schools, the intervention cohort improved an average of 8.2% from pretest to posttest on the overall OSCE score, whereas the comparison cohort improved an average of 4.8% ($P=.006$). Similar differences in change scores were evidenced at each school, and the differences were statistically significant at 2 of the 3 schools. Analyses of change in our other outcome variables yielded similar results. Change score analyses confirmed the findings that, across schools and within each school, improvement in communication skills was significantly greater for students exposed to the intervention than for those who were not.

COMMENT

Our data suggest that dedicated communications curricula significantly improved students' competence in performing skills known to affect outcomes of care. Conduct of a multisite study with comparable interventions but varied curricular implementation at each school permitted a powerful evaluation of the effect of communications training. The commonalities in the in-

Figure. Mean Change in Overall Communication Score for Intervention Cohort vs Comparison Cohort



OSCE indicates objective structured clinical examination. Error bars indicate 95% confidence intervals.

Table 4. Effect of the Communications Intervention: Posttest OSCE Scores Adjusted by Pretest Scores (N = 293)

	Posttest OSCE Scores			P Value by t Test
	Comparison Cohort (n = 138)	Intervention Cohort (n = 155)	Mean Difference (95% CI)*	
Checklist scores, % (SD)				
Overall OSCE†	60.4 (5.1)	65.4 (5.4)	5.1 (3.9 to 6.3)	<.001
Relationship development and maintenance†	57.7 (5.6)	62.9 (6.9)	5.3 (3.8 to 6.7)	<.001
Organization and time management†	74.6 (4.0)	76.4 (3.5)	1.8 (1.0 to 2.7)	<.001
Checklist scores by case type, % (SD)				
Assessment of patient's problem and situation†	60.4 (4.1)	67.3 (4.2)	6.9 (5.9 to 7.8)	<.001
Patient education and counseling	63.3 (7.0)	64.1 (7.2)	0.8 (-0.8 to 2.5)	.31
Negotiation and shared decision making†	58.2 (5.0)	63.9 (5.2)	5.7 (4.5 to 6.9)	<.001
Global recommendation rating (SD)†	3.09 (0.19)	3.17 (0.17)	0.08 (0.04 to 0.12)	<.001

Abbreviations: CI, confidence interval; OSCE, objective structured clinical examination.

*Rounding errors apply to differences.

†Main effect of cohort was significant by analysis of variance. Multivariate analysis of variance established a significant effect of cohort for all scores combined.

terventions across the schools include use of a well-documented model of teaching that emphasizes experiential modes and adult learning principles; dedicated time to teach and practice communications skills; consensus on an inclusive core of skills empirically linked with positive patient outcomes; integrated application of these skills with clinical work during third-year required clerkships in several different medical disciplines; formal faculty development to ensure effective guidance and feedback to students on their performance; teaching strategies that afford application and practice of skills in varying clinical contexts and patient problems; and pacing and feedback tailored to the strengths and deficits of individual students.

Differences in the interventions among the schools are also informative in assessing the generalizability of these findings. While these differences may account for some of the variation in the magnitude of the impact observed among the schools, they also suggest the degree of latitude in approaches that may be expected to yield a favorable effect. The 3 schools differed considerably in the nature of teaching communication skills prior to the advent of the initiative. One school had a well-equipped skills laboratory for assessing student interaction with SPs while the others had more limited commitment to supervised, experiential learning. The schools also differed in the extent to which communications teaching prior

to the initiative was concentrated largely in discrete blocks (eg, courses on interviewing skills during the pre-clinical years) or diffused more extensively throughout the curriculum. Student receptivity to communications training likely differed. In mounting the intervention, individual clerkship directors at one of the schools initiated plans for introducing coverage of specific communications skills in their rotations and took responsibility for implementation, while direction was more centralized at the others. The heterogeneity in the baseline status of the 3 schools with regard to communications teaching is also instructive; in all cases, superimposition of the common elements of the initiative culminated in significant

Table 5. Impact of the Communications Intervention by School: Overall Posttest OSCE Scores Adjusted by Pretest Scores

	Comparison Cohort (n = 63)	Intervention Cohort (n = 65)	Mean Difference (95% CI)	P Value by t Test
School A				
Checklist scores, % (SD)				
Overall OSCE*	60.8 (5.3)	65.5 (5.3)	4.7 (2.8 to 6.5)	<.001
Relationship development and maintenance*	57.4 (5.0)	65.0 (5.7)	7.5 (5.6 to 9.4)	<.001
Organization and time management	75.4 (3.1)	77.2 (2.7)	1.8 (0.9 to 2.9)	<.001
Checklist scores by case type, % (SD)†				
Assessment of patient's problems and situation*	62.2 (4.4)	66.9 (3.6)	4.8 (3.4 to 6.2)	<.001
Patient education and counseling*	64.7 (6.1)	64.3 (6.9)	-0.4 (-1.9 to 2.5)	.73
Negotiation and shared decision making*	57.0 (4.3)	64.9 (5.1)	7.9 (6.3 to 9.5)	<.001
Global recommendation rating (SD)	3.11 (0.20)	3.17 (0.20)	0.06 (-0.01 to 0.13)	.07
School B				
Checklist scores, % (SD)				
Overall OSCE*	59.3 (5.1)	62.3 (4.3)	3.1 (0.9 to 5.2)	<.001
Relationship development and maintenance*	56.1 (6.9)	56.7 (6.6)	0.7 (-2.4 to 0.4)	.67
Organization and time management	71.4 (3.5)	74.0 (3.0)	2.6 (1.2 to 4.1)	<.001
Checklist scores by case type, % (SD)†				
Assessment of patient's problems and situation*	NA	NA	NA	NA
Patient education and counseling*	NA	NA	NA	NA
Negotiation and shared decision making*	NA	NA	NA	NA
Global recommendation rating (SD)	3.03 (0.18)	3.19 (0.16)	0.15 (0.08 to 0.23)	<.001
School C				
Checklist scores, % (SD)				
Overall OSCE*	60.6 (4.7)	68.0 (5.0)	7.3 (5.2 to 9.4)	<.001
Relationship development and maintenance*	59.5 (4.8)	65.5 (5.2)	6.0 (3.8 to 8.1)	<.001
Organization and time management	76.2 (4.0)	77.3 (3.8)	1.1 (-0.5 to 2.8)	.19
Checklist scores by case type, % (SD)†				
Assessment of patient's problems and situation*	NA	NA	NA	NA
Patient education and counseling*	59.4 (8.3)	66.2 (9.0)	6.9 (3.1 to 10.6)	<.001
Negotiation and shared decision making*	NA	NA	NA	NA
Global recommendation rating (SD)	3.11 (0.17)	3.16 (0.13)	0.05 (-0.02 to 0.11)	.14

Abbreviations: CI, confidence interval; NA, data not applicable; OSCE, objective structured clinical examination.

*Interaction effect of cohort by school was significant by analysis of variance. Multivariate analysis of variance established a significant interaction effect for all scores combined.

†Did not achieve sufficient reliability (Cronbach $\alpha > 0.50$) for inclusion as outcome variable at all schools and was therefore omitted in some instances.

improvement in multiple dimensions of student performance.

The effect sizes revealed in our analyses are in the moderate to large range by statistical norms.⁴⁴ It is unclear, however, how these results would translate into clinical outcomes. The state of knowledge and evidence base as yet does not yield a metric for calculating the clinical significance of a 5% absolute difference in overall OSCE performance. Nonetheless, our performance measures are grounded in evidence of their importance to favorable patient outcomes.⁷ For example, findings from the Medical Outcomes Study demonstrated that small differences in patients' ratings of their physician's participatory decision-making style (a construct that informed the selection of our checklist items assessing negotiation and shared decision making) are associated with marked differences in important patient behaviors (eg, remaining with that physician for 12 months or more).²⁸

Elements of our study design suggest that the effect sizes are a conservative estimate of the actual impact of the intervention. The OSCE was designed during the year prior to the development of the curriculum at each site to assess the comparison cohort of trainees. Thus, the content of the OSCE was not refined so as to address distinctive elements of the interventions. Furthermore, the goals governing design of the OSCE had to be sufficiently broad to apply to all 3 schools. In contrast, an OSCE designed to focus on the specificities in the scope and content of each intervention would likely yield larger effect sizes. A corollary is that one of the strengths of our instrumentation is its apparent ability to detect improvement in performance with regard to key patient care competencies attributable to curricula of varied structure and content. As such, the OSCE may have widespread utility for teaching and evaluating communication in diverse settings.

A randomized trial would have been less subject to bias than the cohort design used in our study. To implement a randomized trial at each school, how-

ever, would have required offering 2 different versions of each of the required clerkships and randomly assigning students to either the traditional or new version of each. The logistical obstacles implicit in this design were beyond the capacity of any of the 3 institutions to surmount. Apart from the challenges of maintaining the alternative versions of each clerkship, accommodation to students' schedules and minimization of contamination among intervention and control groups would have been difficult. Nonetheless, our pretest and posttest assessment of student cohorts in successive years addresses some of the potential biases introduced by nonrandom assignment, and we were able to control for the effect of individual differences in student abilities at baseline.

Implementation of the OSCE at a central site would have improved monitoring and reliability of SP case portrayal and ratings. Alternatively, using the same SPs to administer the OSCE at each of the 3 schools would have had similar benefits. Maintaining reliability among different sets of SPs in 3 cities over the course of 2 years was challenging. However, the cost of transporting medical students to a central location or of dispatching 1 set of SPs to 3 medical schools for 4 cycles of OSCEs was prohibitive. One benefit of these constraints is that our approach has yielded protocols for standardized training, assessment of fidelity of case portrayal, and analysis of reliability of student ratings, all of which might be useful to others in mounting similar OSCEs in varied settings.

In summary, our findings suggest that medical schools that incorporate dedicated communications training in clinical clerkships can expect improved student performance with regard to key patient care competencies. The evaluation effort yielded a comprehensively documented, 10-station communications OSCE that may be useful to other institutions for teaching as well as evaluation purposes.

Author Contributions: *Study concept and design:* Yedidia, Gillespie, Kachur, Schwartz, Ockene, Snyder, Lazare, Lipkin.

Acquisition of data: Yedidia, Gillespie, Kachur, Schwartz, Ockene, Snyder.

Analysis and interpretation of data: Yedidia, Gillespie, Schwartz, Chepaitis, Lipkin.

Drafting of the manuscript: Yedidia.

Critical revision of the manuscript for important intellectual content: Yedidia, Gillespie, Kachur, Schwartz, Ockene, Chepaitis, Snyder, Lazare, Lipkin.

Statistical expertise: Yedidia, Gillespie, Chepaitis.

Obtained funding: Yedidia, Lipkin.

Administrative, technical, or material support: Yedidia, Gillespie, Kachur, Schwartz, Ockene, Chepaitis, Snyder, Lipkin.

Study supervision: Yedidia, Schwartz, Ockene, Snyder, Lazare, Lipkin.

Funding/Support: The project was supported by a grant from the Josiah Macy, Jr Foundation.

Acknowledgment: We acknowledge our collaborators on the Macy Initiative in Health Communication, especially the leadership of the curriculum committee (Adina Kalet, Regina Janick, Michele Pugnaire, Emily Ferrara, Kathy Cole-Kelly, and Susan Wentz), as well as Ted Parran and Sondra Zabar.

REFERENCES

1. Association of American Medical Colleges. *Contemporary Issues in Medicine: Communication in Medicine*. Washington, DC: Association of American Medical Colleges; 1999. Report 3 of the Medical School Objectives Project.
2. Liaison Committee on Medical Education. *Functions and Structure of a Medical School*. Washington, DC: Liaison Committee on Medical Education; 1998.
3. Novack DH, Volk G, Drossman DA, Lipkin M. Medical interviewing and interpersonal skills teaching in US medical schools: progress, problems, and promise. *JAMA*. 1993;269:2101-2105.
4. Makoul G. Essential elements of communication in medical encounters: the Kalamazoo consensus statement. *Acad Med*. 2001;76:390-393.
5. Aspegren K. BEME guide No. 2: teaching and learning communications skills in medicine—a review with quality grading of articles. *Med Teach*. 1999;21:563-570.
6. Boon H, Stewart M. Patient-physician communication assessment instruments: 1986 to 1996 in review. *Patient Educ Couns*. 1998;35:161-176.
7. Epstein RM, Hundert EM. Defining and assessing professional competence. *JAMA*. 2002;287:226-235.
8. Kraan HF, Crijnen AA, van der Vleuten CP, Imbos T. Evaluation instruments for medical interviewing skills. In: Lipkin M, Putnam SM, Lazare A, eds. *The Medical Interview: Clinical Care, Education, and Research*. New York, NY: Springer; 1995:460-472.
9. Lipkin M, Kaplan C, Clark W, Novack DH. Teaching medical interviewing: the Lipkin model. In: Lipkin M, Putnam SM, Lazare A, eds. *The Medical Interview: Clinical Care, Education, and Research*. New York, NY: Springer; 1995:422-435.
10. Fallowfield LJ, Lipkin M, Hall A. Teaching senior oncologists communication skills: results from phase I of a comprehensive longitudinal program in the UK. *J Clin Oncol*. 1998;16:1961-1968.
11. Lazare A, Putnam SM, Lipkin M. Three functions of the medical interview. In: Lipkin M, Putnam SM, Lazare A, eds. *The Medical Interview: Clinical Care, Education, and Research*. New York, NY: Springer; 1995:3-19.
12. Kalet A, Cole-Kelly K, Pugnaire M, Janick R, Ferrara E. Curriculum for the Macy Initiative in Health Communication. Available at: <http://endeavor.med.nyu.edu/facDev/model/m00a.html>. Accessed July 11, 2003.
13. Janick R, Kalet A, Zabar S. Faculty development on-line: an observation and feedback module. *Acad Med*. 2002;77:460-461.

14. Stone S, Mazor K, Devaney-O'Neil S, et al. Development and implementation of an objective structured teaching exercise (OSTE) to evaluate improvement in feedback skills following a faculty development workshop. *Teach Learn Med.* 2003;15:7-13.
15. Silverman J, Kurtz S, Draper J. *Skills for Communicating With Patients.* Oxon, England: Radcliffe Medical Press; 1998.
16. Stiles WB, Putnam SM, James SA, Wolf MH. Dimensions of patient and physician roles in medical screening interviews. *Soc Sci Med.* 1979;13A:335-341.
17. Roter DL, Hall JA. Physicians' interviewing styles and medical information obtained from patients. *J Gen Intern Med.* 1987;2:325-329.
18. Tuckett D, Boulton M, Olson C, et al. *Meetings Between Experts: An Approach to Sharing Ideas in Medical Consultations.* London, England: Tavistock; 1985.
19. Maguire P, Falkner A, Booth K, Elliott C, Hillier V. Helping cancer patients disclose their concern. *Eur J Cancer.* 1996;32A:78-81.
20. Eisenthal S, Emery R, Lazare A, Udin H. "Adherence" and the negotiated approach to parenthood. *Arch Gen Psychiatry.* 1979;36:393.
21. Brody DS, Miller SM, Lerman CE, Smith DG, Caputo GC. Patient perception of involvement in medical care: relationship to illness attitudes and outcomes. *J Gen Intern Med.* 1989;4:506-511.
22. Kupst MJ, Dresser K, Schulman JL, Paul MH. Evaluation of methods to improve communication in the physician-patient relationship. *Am J Orthopsychiatry.* 1975;45:420-429.
23. Bertakis KD. The communication of information from physician to patient: a method for increasing patient retention and satisfaction. *J Fam Pract.* 1977;5: 217-222.
24. Kaplan SH, Greenfield S, Ware JE. Assessing the effects of physician-patient interactions on the outcomes of chronic disease. *Med Care.* 1989;27:S110-S127.
25. Stewart MA, Belle Brown J, Weston W, et al. *Patient-Centered Medicine: Transforming the Clinical Method.* Thousand Oaks, Calif: Sage; 1995.
26. Stewart MA, Belle Brown J, Donner A, et al. The impact of patient-centered care on patient outcomes in family practice. London, Ontario: Thames Valley Family Practice Research Unit; 1997.
27. Schulman BA. Active patient orientation and outcomes in hypertensive treatment. *Med Care.* 1979; 17:267-281.
28. Kaplan SH, Greenfield S, Gandek B, Rogers WH, Ware JE. Characteristics of physicians with participatory styles. *Ann Intern Med.* 1996;124:497-504.
29. Beckman HB, Frankel RM. The effect of physician behavior on the collection of data. *Ann Intern Med.* 1984;101:692-696.
30. Rowe MB. Wait time: slowing down may be a way of speeding up. *J Teach Educ.* 1986;37:43-50.
31. Wasserman RC, Inui TS, Barriatua RD, Carter WB, Lippincott P. Pediatric clinicians' support for parents makes a difference: an outcome-based analysis of clinician-parent interaction. *Pediatrics.* 1984;74:1047-1053.
32. Bertakis KD, Roter D, Putnam SM. The relationship of physician medical interview style to patient satisfaction. *J Fam Pract.* 1991;32:175-181.
33. Weinberger M, Greene JY, Mamlin JJ. The impact of clinical encounter events on patient and physician satisfaction. *Soc Sci Med.* 1981;15E:239-244.
34. Levinson W, Roter DL, Mullooly JP, Dull VT, Frankel RM. The relationship among malpractice claims among primary care physicians and surgeons. *JAMA.* 1997;277:553-559.
35. Ewing JA. Detecting alcoholism: the CAGE questionnaire. *JAMA.* 1984;252:1905-1907.
36. Hodges B, Regehr G, McNaughton N, Tiberius R, Hanson M. OSCE checklists do not capture increasing levels of expertise. *Acad Med.* 1999;74:1129-1134.
37. Cohen R, Rothman AI, Poldre P, Ross J. Validity and generalizability of global ratings in an objective structured clinical exam. *Acad Med.* 1991;66:545-548.
38. 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. *Acad Med.* 1999;74:1028-1032.
39. Colliver JA, Swartz MH, Robbs RS, Lofquist M, Cohen D, Verhulst SJ. The effect of using multiple standardized patients on the inter-case reliability of a large-scale standardized patient examination administered over an extended testing period. *Acad Med.* 1998; 73:S81-S83.
40. Florek LM, de Champlain AF. Assessing sources of score variability in a multi-site medical performance assessment: an application of hierarchical linear modeling. *Acad Med.* 2001;76:S93-S95.
41. Pedhazur EJ, Pedhazur-Schmelkin L. *Measurement, Design, and Analysis: An Integrated Approach.* Hillsdale, NJ: Lawrence Erlbaum Associates; 1991.
42. Maxwell SE, Delaney H. *Designing Experiments and Analyzing Data: A Model Comparison.* Belmont, Calif: Wadsworth; 1989.
43. Huitema BE. *The Analysis of Covariance and Alternatives.* New York, NY: John Wiley & Sons; 1980.
44. Cohen J. *Statistical Power Analysis for the Behavioral Sciences.* 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
45. Halpern SD, Karlawish JH, Berlin JA. The continuing unethical conduct of underpowered clinical trials. *JAMA.* 2002;288:358-362.

Many errors, of a truth, consist merely in the application of the wrong names of things.

—Baruch Spinoza (1632-1677)