Correlates of Physicians' Endorsement of the Legalization of Physician-assisted Suicide

KAREN D. NOVIELLI, MOHAMMADREZA HOJAT, THOMAS J. NASCA, JAMES B. ERDMANN, and J. JON VELOSKI

Although most physicians recognize a duty to provide compassionate end-of-life care, they often feel ill prepared to do so. Of particular controversy is physician-assisted suicide. Physician-assisted suicide is commonly defined as the practice of providing a competent patient with a prescription for medication for the patient to use with the primary intention of ending his or her own life. In a recent survey of approximately 2,000 U.S. physicians, 3.3% reported that they had written at least one prescription to hasten death.[1] Eleven percent reported they would write a prescription to hasten death if requested to do so under the current legal system. If legalized, 36% of the physicians would be willing to write a prescription to hasten death.[1]

Consistent with the diversity of physicians' opinions about the practice of assisted suicide, attitudes toward its legalization are also divided. When physicians in Michigan were asked to choose between legalizing or banning assisted suicide, 56% favored legalizing it, while 37% voted for a specific ban.[2]

Several studies have examined the demographic correlates of physicians' attitudes towards assisted suicide. Although age and sex were unrelated to opinions about assisted suicide,[3] race was related. Furthermore, physicians' and patients' preferences for particular approaches to end-of-life care followed similar racial patterns. White physicians were more likely than African American physicians to endorse assisted suicide in terminal care scenarios.[3] Catholic and devoutly religious physicians were also less likely than others to endorse it.[4,5]

Physicians' specialties may also help explain these differences of opinion. Oncologists were more likely to oppose assisted suicide.[6,7] Similarly, support was higher among psychiatrists than among emergency medicine physicians.[4,8,9] Only one study investigated the rationales for physicians' views on assisted suicide. One third of physicians in this study felt that it was immoral, 34% felt that it violated professional ethics, and 30% felt that it conflicted with their own religious beliefs.[10]

Since the legalization of physician-assisted suicide is an area where opinion is sharply divided, research is needed to understand the basis of physicians' beliefs about it. This study was designed to examine the extent and correlates of physicians' endorsements of the legalization of assisted suicide with regard to their specialties, sex, and opinions about certain other contemporary issues in the U.S. health care system.

Method

Graduates of Jefferson Medical College from the classes of 1987–1992 (N = 1,271) who were practicing medicine in the United States comprised the study population. Based on a search of relevant literature and two pilot studies,[11] a survey was developed that consisted of 33 items to be answered on a five-point Likert scale ("strongly agree" = 5, to "strongly disagree" = 1). The survey addressed five aspects of changes in the U.S. health care system influencing medical education, quality of care, patient referral, cost of care, ethical issues, and sociopolitical matters[11] (copies of the survey are available from the authors). The item reading "Physician-assisted suicide should be legalized" was used as the dependent variable in the present study.

The questionnaires were mailed in May 1998, followed by three reminders mailed to non-respondents at three-week intervals. Useable forms were returned by 835 physicians (66% response rate), of whom 830 responded to the item on the legalization of physician-assisted suicide. The respondents included 578 (69%) men and 257 (31%) women, with a mean age of 35.8 years. The specialties of respondents were distributed as follows: family practice, 116 (14%), general internal medicine, 85 (10%), pediatrics, 38 (5%), emergency medicine, 49 (6%), obstetrics–gynecology, 34 (4%), surgery and surgical subspecialties, 47 (6%), psychiatry, 28 (3%), hospital-based specialties (anesthesiology, pathology, and radiology), 97 (12%), medical subspecialties, 86 (10%), and other specialties and subspecialties, 255 (30%). Statistical analysis included bivariate and multivariate correlation, *t* test, chi-square, and *z* test for proportions.

Results

No significant difference was found between respondents and non-respondents with respect to gender (31% versus 33% women, respectively), age (35.8 versus 35.9 years), full-time salaried faculty appointment (14% versus 12%), and primary care practice (which was defined as family medicine, general internal medicine, and general pediatrics) (29% versus 34%).

Similarly, no difference was found for academic performance measures such as scores on the United States Medical Licensing Examinations, Steps 1–3, and clinical competence ratings provided by residency program supervisors at the end of the first postgraduate training year in three competence areas of "data-gathering and processing skills," interpersonal skills and attitudes," and "socioeconomic aspects of patient care."[1,12]

Respondents' Endorsement of Legalization of Physician-assisted Suicide. Of the 830 respondents, 284 (34%) endorsed legalization—73 (9%) "strongly agreed," and 211 (25%) "agreed"; and 340 (41%) opposed it—189 (23%) "disagreed," 151 (18%) "strongly disagreed," and 206 (25%) expressed "no opinion." The response patterns were similar for physicians who graduated in the six different cohorts.

Correlates of Endorsement of Legalization of Assisted Suicide. The endorsement rates for legalization of physician-assisted suicide were examined by the following variables:

- *Demographics.* Endorsement of legalization was unrelated to age and gender. Although the small number of African-American and Hispanic physicians in the sample was insufficient for meaningful statistical analysis. Asian physicians ($n = 48$) were significantly more likely (63%) than were whites ($n = 557$) to endorse (43%) legalization (*z*-test for proportions = 2.85, $p < .01$).
- *Specialty.* Orthopedic surgeons endorsed assisted suicide at the highest rate, which was 52%, followed by psychiatrists (41%), and physicians in the hospital-based specialties (40%). The lowest rates were for medical subspecialists (25%), general internists (28%), emergency medicine physicians (31%), family physicians (33%), and general pediatricians (34%). These differences in attitudes toward legalization among specialties were statistically significant ($\chi^2_{(20)} = 33.7$, $p < .05$).
- *Postgraduate ratings of clinical competence.* The physicians who endorsed legalization had been rated significantly lower by their

TABLE 1. Bivariate and Multiple Correlations and Regression Coefficients Predicting 830 Physicians' Endorsements of Physician-assisted Suicide, Jefferson Medical College*

Predictor†	Bivariate r	Regression Coefficient
Physicians should unionize to maintain the influence of their profession.	.17§	.12‡
The present paradigm of medical education does not take into account the psychosocial factors related to illness.	.15§	.15§
Government should be responsible for regulating policies that influence the quality of care.	.12§	.12‡
Learning to work in a changing health care environment should become an essential part of medical education.	.11§	.14§
Physicians involved in HMOs or other types of managed care order fewer tests than those in private practice.	.11§	.12‡
The future of health care should be based on the needs of society not on the satisfaction of physicians.	−.11§	−.09§
Physicians involved in managed care have the same dedication to their patients as physicians in fee-for-service.	−.08§	−.08‡
Intercept		2.3§
Multiple R		.30§

*Participants were 830 physicians who graduated from Jefferson Medical College between 1987 and 1992.
†Items on 33-item survey that correlated either positively or negatively with respondents' endorsement of physician-assisted suicide.
‡$p < .05$; §$p < .01$.

residency program directors in the postgraduate clinical competence areas of "interpersonal skills and attitudes" ($F_{(1, 452)} = 6.25$, $p < .01$), and "socioeconomic aspects of patient care" ($F_{(1, 452)} = 6.94$, $p < .01$). No significant difference was noted in the area of "data gathering and processing skills."

- *Other significant predictors of endorsement of legalization.* Bivariate correlations between responses to the item on legalization and those for other 32 items in the survey were examined. Nine items had statistically significant correlations with endorsement of legalization. A stepwise multiple regression algorithm was used, in which numerical weights assigned to responses to the item on legalization were considered as the dependent variable (criterion measure) and numerical weights assigned to the nine items of the survey that had significant correlations with responses on the physician-assisted suicide item were the independent variables (predictors). Only seven items contributed significantly ($p < .05$) to the multiple regression model, which is summarized in Table 1. Five contributed positively in that endorsement of legalization of physician-assisted suicide was associated with agreement with those items. Two contributed negatively, meaning that endorsement of legalization was associated with disagreement with those items.

As reported in Table 1, those who endorsed legalization were more likely to agree that physicians should unionize ($r = .17$, $p < .01$), that the present paradigm of medical education does not take into account the psychosocial factors related to illness ($r = .15$, $p < .01$), that government should take responsibility to regulate health care policies ($r = .12$, $p < .01$), that learning to work in a changing health care environment should become an essential part of medical education ($r = .11$, $p < .01$), and that physicians who work with managed care organizations order fewer tests than their counterparts in private practice ($r = .11$, $p < .01$).

Conversely, the physicians who endorsed legalization were more likely to disagree that the future of health care should be based on the needs of society rather than on physicians' satisfaction ($r = -.11$, $p < .01$) and that physicians in HMOs as compared with those in other settings have similar dedication to their patients ($r = -.08$, $p < .05$). The multivariate correlation was .30, $p < .01$ (see Table 1).

It is noteworthy that the responses to the item on legalization were not correlated with several other items, including the consideration of cost as an important factor in patient care decisions, physicians' support for the efforts of government to ration care, and the role that organized medicine should take with respect to social issues that can influence the well-being of society.

Discussion, Conclusions, and Implications

The findings of the present study support prior research showing that physicians hold widely disparate views regarding the legalization of physician-assisted suicide. More physicians in our study were opposed to legalization (41%) than supported it (34%), and a significant fraction of these physicians (25%) had not formed an opinion. The proportion of physicians in our study favoring legalization was similar to those in other survey work in this area.[2] Almost all respondents endorsed medical school preparation for, and subsequent provision of, compassionate care at the end of life (92%), suggesting that the differences of opinion related only to the controversial area of assisted suicide and not to caring for the dying patient in general.

Our study found that physicians in the people-oriented specialties most associated with direct and ongoing patient contact that included treatment of dying patients (general medicine, family medicine, and medical subspecialties) were less likely to endorse legalization than were technology-oriented physicians, including hospital-based specialists and orthopedic surgeons. Experience with the first year of legalized physician-assisted suicide in Oregon acknowledges the great emotional toll on physicians directly involved in its implementation.[14] The emotional burden and the acknowledged complexities in caring for dying patients may make physicians involved in this process more reluctant to endorse legalization. An interesting corollary suggested by our findings is that physicians endorsing legalization were less comfortable with their medical school training in the psychosocial aspects of care and were rated poorer in the areas of interpersonal skills and attitudes and in socioeconomic aspects of patient care in the first year of residency.

It is not known to what degree opinions about legalization are subject to modification by educational experiences during medical school. A recent study that examined medical students' views on physician-assisted suicide found that fourth-year medical students in Oregon were less likely than were fourth-year medical students in other areas of the country to be willing to provide a patient with a lethal prescription.[15] The authors suggested that a change in willingness to comply with legalized physician-assisted suicide might have occurred as a result of experience with such requests from dying patients.

Unlike many areas of medical education where knowledge is largely dependent on didactic teaching, care of the dying and attitudes towards assisted suicide are likely to be influenced primarily by personal experiences as well as the moral, ethical, and political tenets that adults bring to medical training. In addition to explor-

ing more closely the relationship between these personal beliefs and attitudes, an important priority for research is to determine whether attitudes towards care of the dying and physician-assisted suicide could be modified by education. Evaluation of the impact of educational experiences such as structured exposure to palliative care or rotations in a hospice service for medical students and residents would help to answer these questions. As the U.S. health care system moves from theory to practice regarding physician-assisted suicide, more research is needed to explore further the impact of legalization on physicians and their patients.

The advantages of this survey include the large sample size, gender composition, and specialty and geographic distribution of the participants that represent a broad spectrum of the population of physicians. Despite these advantages, one limitation of our study is that it ascertained physicians' views of the legalization of assisted suicide rather than their views of its practice. However, the two concepts seem logically related. The primary purpose of the survey was to gather views of multiple issues in the current health care system, including attitudes toward legalization of assisted suicide. Another limitation is that the results of this study of young physicians who graduated from a single private medical school in the Northeast may not be fully generalizable to all U.S. physicians. However, the distribution of reactions is similar to that reported in the literature.[2]

As physicians hold an influential position in the public debate on the legalization and practice of physician-assisted suicide, it is important to further understand the bases for their strong and disparate views. Futher research in this area should elucidate the political, moral, and ethical frameworks that physicians bring to this topic. Specifically, it is essential to understand the degree to which physicians' views on the legalization of physician-assisted suicide are subject to modification by medical education in general,[16] and by experiences with dying patients in particular.

Development of the foundation for this study was supported, in part, by a grant from the Bureau of Health Professions, Health Resources and Services Administration, USDHHS, under Cooperative Agreement 1 U76 MB00002-03, Centers for Medical Education Research and Policy.

Correspondence: Karen Novielli, MD, Department of Family Medicine, Jefferson Medical College, 401 Curtis, Philadelphia, PA 19107; e-mail: ⟨karen.novielli@mail.tju.edu⟩.

References

1. Meier DE, Emmons CA, Wallenstein S, et al. A national survey of physician-assisted suicide and euthanasia in the United States. N Engl J Med. 1998;338:1193–201.
2. Bachman JG, Alcser KH, Doukas DJ, et al. Attitudes of Michigan physicians and the public toward legalizing physician-assisted suicide and voluntary euthanasia. N Engl J Med. 1996;334:303–9.
3. Mebane EW, Oman RF, Kroonen LT, Goldstein MK. The influence of physician race, age and gender on physician attitudes toward advanced care directives and preferences for end-of-life decision-making. J Am Geriat Soc. 1999;47:579–91.
4. Schmidt TA, Zechnich AD, Tilden VP, et al. Oregon emergency physicians' experiences with, attitudes toward, and concerns about PAS. Acad Emerg Med. 1996;3:938–45.
5. Siaw LK, Tan SY. How Hawaii's doctors feel about physician-assisted suicide and euthanasia: an overview. Hawaii Med J. 1996;55:296–8.
6. Abramson N, Stokes J, Weinreb NJ, Clark WS. Euthanasia and doctor-assisted suicide: responses by oncologists and non-oncologists. South Med J. 1998;91:637–42.
7. Cohen JS, Fihn SD, Boyko, EJ, Jonsen AR, Wood RW. Attitudes toward assisted suicide and euthanasia among physicians in Washington State. N Engl J Med. 1994;331:89–94.
8. Ganzini L, Fenn DS, Lee MA, Heintz RT, Bloom JD. Attitudes of Oregon psychiatrists toward physician-assisted suicide. Am J Psychiatry. 1996;153:1469–75.
9. Roberts LW, Roberts BB, Warner TD, et al. Internal medicine, psychiatry, and emergency medicine residents' views of assisted death practices. Arch Intern Med. 1997;157:1603–9.
10. Lee MA, Nelson HD, Tilden VP, Ganzine L, Schmidt TA, Tolle SW. Legalizing assisted suicide—views of physicians in Oregon. N Engl J Med. 1996;331:310–5.
11. Hojat M, Veloski JJ, Louis DZ, et al. Perceptions of medical school seniors of the current changes in the U.S. health care system. Eval Health Prof. 1999;22:169–83.
12. Hojat M, Veloski JJ, Borenstein BD. Components of clinical competence ratings: an empirical approach. Educ Psychol Meas. 1986;46:761–9.
13. Hojat M, Borenstein BD, Veloski JJ. Cognitive and noncognitive factors in predicting the clinical performance of medical school graduates. J Med Educ. 1988;63:323–5.
14. Chin AE, Hedberg K, Higginson GK, Fleming DW. Legalized physician-assisted suicide in Oregon—the first year's experience. N Engl J Med. 1999;340:577–83.
15. Mangus RS, Dipiero A, Hawkins CE. Medical students' attitudes toward physician-assisted suicide. JAMA. 1999;282:2080–1.
16. Barzansky B, Veloski JJ, Miller R, Jonas HS. Education in end-of-life care during medical school and residency training. Acad Med. 1999;74(10 suppl):S102–S104.

Learning Adolescent Psychosocial Interviewing Using Simulated Patients

KIM BLAKE, KAREN V. MANN, DAVID M. KAUFMAN, and MURRAY KAPPELMAN

The area of communication skills in adolescent medicine is emerging as a distinct and important part of the undergraduate curriculum. An appropriate level of confidence in dealing with the adolescent population is deemed a necessary educational requirement.[1] Skills in psychosocial communication with adolescents differ from those required for younger patients and adults[2–4]; they include discussing confidentiality and adolescent risk-taking activities. Simulated patients can be used effectively in teaching and evaluating of communication skills.[5–6] However, there is no report of using adolescent simulated patients to teach communication skills.

The evidence available is inconclusive regarding the teaching time required to promote retention of communication skills, although a recent review[7] suggests that one day's training or less is not effective. Long-term retention of these skills has been supported by only one paper,[8] suggesting a need to follow students over time to ascertain the effect of communication skills training.

Our study addressed two questions: (1) does feedback from a simulated adolescent patient and simulated mother lead to improvements in fourth-year medical students' psychosocial interviewing of adolescent patients? and (2) does this skill persist following the intervention?

Method

Final-year medical students (N = 68) from March 1998 through May 1999 were invited to participate, and 57 agreed. The 11 who were unavailable to participate were either interviewing for their postgraduate education, involved in presenting their own research, or unable to make the scheduled times for the simulations. Thirty five other class members were either randomly or self-selected to go to offsite locations for pediatrics, and therefore could not participate; however, this group acted as a non-randomized control arm to the study. A two-group (57 students in the intervention group and 35 in the control) prospective randomized double-bind study design was employed. The students were completing an eight-week core pediatrics rotation in a tertiary center, with seven to nine students per rotation.

Study Question 1

Intervention. Four simulated cases were developed, each comprising both a medical component (epilepsy, diabetes, attention deficit disorder, or asthma) and risk-taking activities (smoking, drugs, boyfriend issues) in which the adolescent was scripted to be involved. Nine simulated mothers and ten female adolescents (mean age 13.6 years) were recruited using established procedures.[9] Mother-and-daughter pairs were selected as this is the commonest adolescent presentation in medical practice. Young adolescents were chosen to provide a realistic presentation of this age group, which often presents a challenge to young doctors. The training for standardized feedback was achieved when all mothers reviewed a single taped scenario, scored this independently using a structured form, and then discussed the feedback they would provide the student in a group setting. The adolescents were guided by their partner mothers to give feedback, which the adolescents discussed in a focus group.

At study entry, all students signed informed consent forms. They then interviewed a simulated mother–daughter pair. The students were randomly assigned to receive immediate feedback following the pretest interview from the simulated pair (F^2), or to receive no feedback (F^1). All students conducted a second interview four weeks later using a different case scenario. All students (F^1 and F^2) received feedback from the simulated pair following this post-test interview. Feedback was structured using a written modified Calgary–Cambridge guide[10] and given verbally; both interview content and process were addressed.

Measures. Three measures were taken:

1. Questionnaire. At study entry, demographic data and students' self-ratings of prior experiences with adolescent medicine, confidence in dealing with adolescent patients, and anticipated future work with adolescents were collected.
2. Pre-test. Students conducted a one-hour videotaped interview with a simulated adolescent and mother, using one of the four case scenarios, at the midpoint of their rotation. The videotaped interviews were scored by a psychologist who had been trained to reach an acceptable level of agreement with the principal investigator (KB) using the modified Calgary–Cambridge guide.[10]
3. Post-test. Four weeks later, each student conducted a second videotaped interview, using a different case scenario. Scoring was completed in the same manner as for the pre-test.

Study Question 2

Intervention. The entire final-year class participated in a mandatory ten-station OSCE prior to graduation. This was two to 12 months after participation in the study (mean 6.6 months). One pediatrics station of this OSCE tested general pediatrics knowledge (students' performances in asking about medical aspects of the case) and adolescent psychosocial interviewing (students' performances in asking about psychosocial aspects, e.g., boyfriend, alcohol, drugs). The OSCE included 35 off-site students, those not involved in the adolescent interviewing study, i.e., those who had not been videotaped and had received no feedback (F^0) and 45 of the 57 students who had completed their pediatrics rotation at the tertiary center and who had participated in the study (F^1 and F^2).

Measures. The knowledge score and the psychosocial interviewing score on the pediatrics OSCE station were obtained from the checklists completed by the faculty examiner at the station.

Data Analysis

Study Question 1. A single psychologist, blinded to student group or time of interview, scored the tapes, using a modified Calgary–Cambridge Observation Guide.[10] The psychologist evaluated eight aspects of the encounter: how the student initiated the session, collected information, gathered information, asked the parent for time alone with the patient, dealt with the adolescent alone, and acted before and during the examination and closure. Each section yielded a global score. Within seven of the sections there were between three and ten individual items. The section used to rate when the student was alone with the adolescent included 14 psychosocial elements (i.e., boyfriend issues, smoking, and drugs).

The psychologist derived eight global ratings for each videotape. The global ratings for F^1 and F^2 students at pre- and post-test were compared using a paired t test. Regression analysis was conducted

using student global ratings from the eight sections of the modified Calgary–Cambridge Observaton Guide as the dependent (outcome) variable. The independent (predictor) variables were feedback, case type and simulator, gender, previous medical experience with adolescents, comfort level in relating to adolescents, future career plans, and the students' scores on the pre-test case.

Study Question 2. The knowledge score and the psychosocial interviewing score on the pediatrics OSCE station were compared among the three groups (F^0, F^1, F^2).

Results

Complete data were available for 52 of the 57 students (F^2 = 31; F^1 = 21) who completed both pre- and post-test interviews. Two tapes could not be rated, and three students did not complete the second interview.

Study Question 1. The mean pre-test scores of the group receiving feedback after their first interview (F^2) and those receiving no initial feedback (F^1) were not statistically different (72.93, SD = 9.43 versus 72.77, SD = 8.08; $p = 0.95$). However, the group that received feedback immediately after their first interviews (F^2) scored significantly higher on the post-test (82.81, (SD = 9.79) than did the F^1 group (76.34, SD = 9.43); $p = 0.02$). No significant improvement was seen from pre-test to post-test for the group receiving no initial feedback (F^1). However, the group receiving feedback (F^2) improved significantly from pre-test to post-test (p = 0.02).

Regression analysis revealed that receiving feedback was the only significant predictor ($p = .021$) of students' performances on the post-test case ($R^2 = .10$ for the complete model). The other independent variables did not significantly predict post-test performance. Analyses also were conducted to determine whether or not the particular case scenario used had a significant influence on student performance. No statistically significant influence due to case difference emerged.

Study Question 2. All students participating in the study received feedback either once (F^1) or twice (F^2). Both groups ($n = 45$) had significantly higher mean scores ($p = .023$) on the adolescent psychosocial inquiry on the final-year OSCE station (68.06, SD = 24.07) compared with the students ($n = 35$) who completed their core pediatrics rotation at the offsite placements (F^0) (55.71, SD = 23.16). The groups did not differ significantly ($p = 0.40$) in their mean scores for the general knowledge aspects of this OSCE station (F^1 and F^2) (70.71, SD = 16.88) compared with F^0 (67.53, SD = 16.69).

After the OSCE the students were asked to comment on their clerkship experience. The simulated adolescent encounters were rated as one of the most positive learning experiences in the two years of clerkship.

Discussion

The main study finding is that the important communication skill of interviewing the adolescent patient can successfully be taught to undergraduate medical students. The teaching becomes faculty-independent when the simulated patients are scripted and trained in giving structured feedback. The training period, which was a one-hour interview (experimental), followed by 20 minutes of feedback, was much less than one day, which is the time reported in the literature as necessary for effective learning of these skills. This study poses questions for further research regarding optimal training time and the best method of reinforcement. For psychosocial interviewing with sensitive questioning, clerkship seems the optimal point of instruction; however, there is little evidence to inform where training in communication with adolescents should be placed in the medical curriculum.

There are several limitations to this study. First, the sample was small, although representative of other randomized controlled trials in this field. Second, selection bias may have occurred, as the students who chose to complete their core pediatrics rotations at off-site placements were either randomly or self-selected. However, all students received the core pediatrics tutorials from the tertiary center by teleconference, along with detailed objectives. This ensured that all students received the same didactic curriculum. Third, although the study would have benefited from two independent raters, the increased cost was prohibitive. The psychologist rater was trained to use the modified Calgary–Cambridge Guide[10] and underwent a mid-study validation of his scoring. Fourth, our sample was confined to mothers and daughters; whether the results would differ with mother–son simulator pairs is unclear. Fifth, although this study provides some indication that students' psychosocial communication skills can be improved and maintained over time, follow up was less than a year. Continued tracking of these doctors would be important to see whether this mastery is maintained into the residency years. Finally, application of these results must consider resources. At our medical school, standardized patients frequently supplement current teaching activities, and are part of the diagnostic assessment of student skills throughout the medical school curriculum. Expertise to train and administer such a program is quite involved from a logistic and monetary standpoint; although available at our medical school, this may not be the case everywhere. As this educational initiative relies on a realistic portrayal and structured feedback from the adolescent, time spent in recruitment and training of the standardized patients is important.

Students overwhelmingly commented that feedback from a "real" adolescent was very helpful, as they had received little training in this area. Many of the students were very apprehensive on entry into the study, but were resoundingly positive after they had completed it.

Because of the changing nature of the hospitalized patient population, standardized patients could be used to ensure that each student has exposure to common ambulatory problems. They could help ensure uniformity in teaching and learning of basic clinical skills. Interviewing an adolescent standardized patient who is involved in risk-taking activities provides the student an opportunity to practice psychosocial interviewing in a safe setting. The immediate feedback provided by the adolescent and mother is a powerful teaching tool. The student can then return to the clinical setting to apply these newly acquired skills.

In conclusion, this randomized controlled trial has shown that final-year medical students can be taught adolescent interviewing skills and that these skills are retained for as long as a year. The teaching time required for such an intervention is short (90 minutes), and teaching can be independent of faculty once the simulators' training is completed. As the skill of talking to adolescents and their parents is an important part of physician training, we would recommend that medical schools consider this structured training for their curricula.

This research was part of Dr. Blake's AAMC Fellowship in Medical Education (1997–1998). The authors thank the members of the LRC—Nancy Ruedy, Ruth Partridge, and Susan Wakefield—for the training of simulated patients, and Marilyn Massie-Clarke for statistical analysis. Supported by a grant from the Division of Medical Education, Dalhousie University.

Correspondence: Kim Blake, MD, Dalhousie University, Department of Pediatrics, Dalhousie University 5850/5859 University Avenue, Halifax, NS, Canada B3J 3G9; e-mail: ⟨kblake@is.dal.ca⟩.

References

1. Olson, Al, Kaufman NM, Marshall SG, Woodhead J. General Pediatric Clerkship Curriculum and Resource Manual. Ambulatory Pediatric Association and the Council on Medical Student Education in Pediatrics, Chapel Hill, NC, 1995.

2. Boekeloo B, Schamus L, Cheng T, Simmens S. Young adolescents' comfort with discussion about sexual problems with their physicians. Arch Pediatr Adolesc Med. 1996;150:146–52.
3. Joffe A, Radius S, Gall M. Health counseling for adolescents: what they want, what they get, and who gives it. Pediatrics. 1998;82:481–5.
4. Wissow LS, Roter DL, Wilson ME. Pediatrician interview style and mothers' disclosure of psychological issues. Pediatrics. 1994;93:289–95.
5. Stillman PL, Regan MB, Philbin M, Haley HL. Results of a survey on the use of standardized patients to teach and evaluate clinical skills. Acad Med. 1990;65:288–92.
6. Sanson-Fisher RW, Poole AD. Simulated patients and the assessment of medical students' interpersonal skills. J Med Educ. 1980;14:249–53.
7. Aspegren K. BEME Guide No. 2: Teaching and learning communication skills in medicine—reviewing with quality of articles. Med Educ. 1999;21:563–70.
8. Maguire P, Fairbairn, Fletcher C. Consultation skills of young doctors: benefits of feedback training in interviewing as students persist. BMJ. 1986;292:1573.
9. Blake K, Greaven S. Recruiting and following adolescent standardized patients. Acad Med. 1999;74:584.
10. Kurtz SM, Silverman JD. The Calgary–Cambridge Observation Guides. Med Educ. 1996;30:83–9.

Have Clinical Teaching Effectiveness Ratings Changed with the Medical College of Wisconsin's Entry into the Health Care Marketplace?

DAWN BRAGG, ROBERT TREAT, and DEBORAH E. SIMPSON

Medical schools, as competitors in today's health care marketplace, have the challenge of training future physicians while increasingly relying on clinical revenues.[1] Is teaching compatible with competitive managed care in the future of health care?[2]

Skeff, Bowen, and Irby argue that teaching takes time and that its values must be re-emphasized as a core mission of medical schools.[3] Medical education researchers have reported diminishing amounts of time available for physicians' educational responsibilities to both residents[4] and medical students.[5] Student evaluations reveal that there has been less time available for them in more recent years.[6] Thus, time impacts on education have been documented, but the critical issue to be investigated is whether the quality of teaching has been compromised.

As a large, private medical school, the Medical College of Wisconsin (MCW) has not escaped the grasp of today's competitive health care environment. On December 31, 1995, the John L. Doyne Hospital (JLDH), formerly Milwaukee County General Hospital, was closed. While this facility (a primary practice and clinical teaching site) was purchased by a private adult not-for-profit hospital, it's sale nonetheless serves as a major demarcation point in MCW's transition into today's health care marketplace. Indigent care was now provided on a competitive contract basis. Our faculty formed a clinical practice group to enhance their competitive position in this evolving health care environment. Declining federal support for graduate medical education led to decreased positions in selected specialties and their associated support of medical student education. While the multi-dimensional impact of these changes on medical education, at MCW and elsewhere, will take years to analyze,[7] preliminary analysis can reveal whether the quality of clinical teaching has changed during this time period. This study, therefore, examined whether there have been changes in clinical teaching effectiveness ratings as clinicians at MCW compete for patients and revenue.

Method

The study utilized student ratings of clinical teachers from a longitudinal clinical teaching database implemented in 1992. A standard clinical teaching instrument[8] is used across participating clinical departments. The instrument contains 16 characteristics of effective clinical teaching, derived from a comprehensive review of the literature, rated using a five-point Likert scale (1 = most positive). Items address faculty interaction with students (e.g., actively involved me with patients, provided timely, constructive feedback without belittling me), ability to communicate (e.g., clear, organized, answered my questions clearly), and overall teaching effectiveness. The form is highly reliable, with a coefficient alpha of .96.

Since 1992, third-year medical students have evaluated 295 full-time clinical teachers in pediatrics, internal medicine, family medicine, anesthesiology, and general surgery. For purposes of this study, the data were divided into three time periods, using 1995 as the benchmark date for MCW's entry into health care marketplace: before-entry, 1993–94; at-entry, 1995–96; and after-entry, 1997–98 (numbers of evaluations per period = 1,327, 4,354, and 6,577 respectively).

A three-stage analytic process was used to determine whether students' ratings of clinical teaching had changed during the study period. First, the 16 clinical teaching instrument items were clustered to facilitate analysis using agglomerative hierarchical cluster analysis (HCA).[9] This method has been successfully used to cluster items on standardized tests into psychological dimensions.[10] In HCA for an n-item test, there are n solutions or clusters. In the first step, each item comprises one cluster. At subsequent steps, the procedure combines two clusters from the previous step, based upon the proximity or similarity among each possible pairing of the clusters. The smaller the proximity value, the more similar the two clusters are believed to be. The final cluster, the nth cluster, places all of the items into one cluster. By examining the two- or three-cluster solution for interpretability, a researcher can get a nonparametric perspective on groups of items that may be considered to be dimensionally distinct. Unlike factor analysis, cluster analysis is nonparametric and is a quick way to identify possible dimensions that may exist. In this study, selected clusters of clinical teaching skills were examined for internal consistency using coefficient alpha.

Using these clusters, two-way analysis of variance was performed comparing the cluster means to determine whether (1) students' ratings varied by time period; (2) students' ratings varied by item cluster; and (3) there was an interaction effect between time periods and clusters. Individual items that had been closely associated with the availability of teaching time in previous studies were then analyzed using a one-way analysis of variance to examine differences in student ratings across the three time periods.

Results

A three-cluster solution resulting from the HCA was selected for statistical and substantive reasons and to increase comparability of results with findings from prior factor-analytic studies. Ullian et al.,[11] in their synthesis of factor-analytic studies, reported that while there are varying numbers of factors, most studies suggest four solutions. The three-cluster solution was selected for this study as the two-cluster solution contained many items that did not seem to fit qualitatively and other cluster solutions contained at least one group with fewer than four items, posing a threat to internal consistency. The three clusters were examined qualitatively to assess content validity and their relationship to Ullian's four factors.

The first cluster of clinical teaching skills was labeled *supervisor/person* and contained seven items: supportive of me/had rapport with me, approachable/available, actively involved me with patients, communicated expectations, demonstrated skills/procedures to be learned, provided opportunities to practice diagnostic/assessment skills, and provided feedback without belittling me. The second cluster was labeled *physician/teacher* and contained five items: answered questions clearly, asked questions clearly, explained basis for decisions/actions, clear/organized, and clinically competent/knowledgeable. The third group, containing four items, was labeled *instructor/leader*: took advantage of teaching opportunities, enthusiastic/stimulating, responded to student-initiated learning issues, and emphasized comprehension rather than factual recall. All three item clusters, supervisor/person, physician/teacher, and instructor/leader, were found to be highly reliable (coefficient alpha = .90, .86, .80, respectively). According to Ullian et al., these three clus-

ters define the roles that clinical teachers assume in their interactions with students.

The students' ratings ranged from a minimum of 1 (most positive) to a maximum of 5 (least positive). Mean ratings across the three time periods were found to differ significantly ($p < .001$) (see Table 1). Post-hoc comparisons (i.e., Tukey test) revealed that the mean ratings for the periods were significantly different (all comparisons $p < .001$). Mean student ratings for the three clusters were also significantly different ($p < .001$). Throughout the before-entry, at-entry, and after-entry periods, physician/teacher skills were rated best by third-year students, while supervisor/person skills received the worst ratings (see Table 1). The analysis also showed an interaction between the time periods and the three groups ($p < .001$).

Mean student ratings for the three sets of skills started out positively in the first, before-entry year (see Figure 1). This was due to the fact that in 1993 faculty began to receive the first results of their clinical teaching evaluations. As reported in a prior study, when faculty receive clinical teaching evaluation results, their clinical teaching ratings improve as they immediately seek to address deficits.[12] Mean ratings for supervisor/person and instructor/leader skills increased (became worse) sharply in the second year. Mean ratings for physician/teacher continued to improve throughout the before-entry years. During the at-entry period, mean ratings for supervisor/person and instructor/leader skills continued to increase (becoming worse), but the ratings increased only gradually for physician/teacher. Supervisor/person skills peaked in 1996, the year the faculty practice plan was implemented. Mean ratings for instructor/leader and physician/teacher leveled off between 1995 and 1996. The after-entry period saw improved ratings for the three item clusters. However, none of the cluster ratings returned to the before-entry baseline level.

Of particular importance were the significant differences across time periods among the mean ratings of those characteristics associated with the availability of time. For example, mean ratings of items within the supervisor/person (e.g., supportive of me, approachable/available, actively involved me with patients) followed the increased cluster ratings. However, the ratings for "provided timely, constructive feedback without belittling me," received increasingly poor ratings across the three time periods. Analysis indicated that all four questions within this cluster were significantly different across the time periods ($p < .005$).

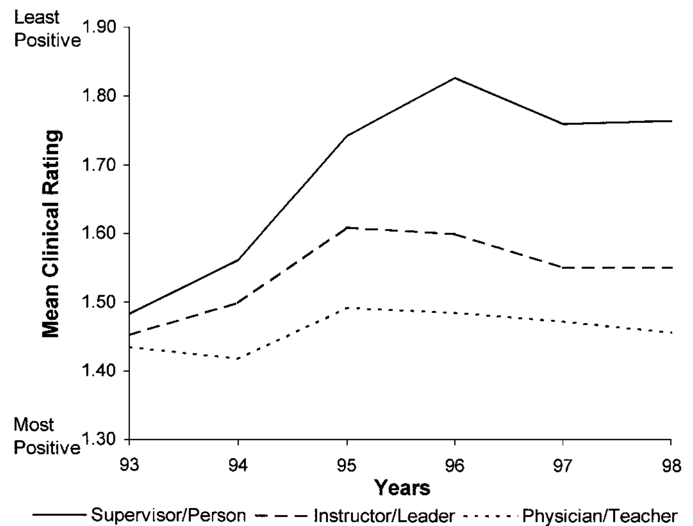

Figure 1. **Students' mean ratings of physicians' clinical teaching skills across the before-entry (1993–94), at-entry (1995–96), and after-entry (1997–98) time periods.**

Table 1. Third-year Medical Students' Ratings of Physicians' Clinical Teaching Skills before, at, and after the Medical College of Wisconsin's Entry into the Health Care Marketplace, 1992–1998

	Rating*		
	Before Entry (n = 1,327)†	At Entry (n = 4,354)†	After Entry (n = 6,577)†
	Mean (SD)	Mean (SD)	Mean (SD)
Skills clusters			
Supervisor/person	1.53 (.57)	1.79 (.72)	1.76 (.70)
Physician/teacher	1.43 (.49)	1.49 (.52)	1.47 (.54)
Instructor/leader	1.48 (.54)	1.60 (.65)	1.55 (.64)
Individual items			
Supportive of me/had rapport with me	1.44 (.71)	1.69 (.90)	1.67 (.88)
Actively involved me with patients	1.37 (.67)	1.75 (.89)	1.67 (.87)
Approachable/available	1.40 (.71)	1.59 (.84)	1.57 (.86)
Provided timely constructive feedback without belittling me	1.58 (.78)	1.76 (.92)	1.80 (.93)

*Scale: 1 = most positive to 5 = least positive.
†n = number of evaluations.

Discussion

Longitudinal analysis of a clinical teaching evaluation data set reveals that the overall effectiveness of our clinical teaching decreased from a before-entry high at the time of entry in the health care marketplace. Over the at-entry study period, evaluations did gradually improve, but did not return to the before-entry baseline rate. However, not all item ratings were equally affected, with physician/teacher skills (e.g., clear/organized, clinically competent) showing the least change and supervisor/person skills (e.g., approachable, available, supportive of me, actively involved me with patients, provided timely, constructive feedback without belittling me) showing the largest decline. The supervisor/person skills, containing the interpersonal items, appear to have been the most profoundly affected by the entry into the health care marketplace.

Although it may be possible that students become more discriminating in their assessments of teaching and teachers over time, this study does not report ratings by the same students over time. This study used ratings by individual third-year classes for six years. In addition, student ratings were averaged over two years for each time period, thus minimizing huge class differences.

HCFA guidelines, increased pressures for clinical productivity, and accountability for cost-effective patient care have led physicians to repeatedly report that they have less time for clinical teaching. The results of this study suggest that there has also been a change in the *quality* of clinical teaching, as measured by the clinical teaching effectiveness ratings over this critical time period, a relationship requiring further study to determine causality. While it is promising that the rating results do appear to have improved following an initial decline during the at-entry period, the fact that these ratings did not return to baseline levels is distressing.

Supervisor/person skills are critical components of the teaching/learning process, as education is enhanced when there is a supportive relationship between the learner and the teacher.[13] Medical schools must prepare clinical educators with teaching skills that are effective and efficient in today's time-pressured clinical environments and implement real reward structures that recognize the value of time spent in clinical teaching if we are to maintain the quality of our clinical education.

Correspondence: Dawn Bragg, PhD, Assistant Director, Office of Educational Services, Medical College of Wisconsin, 8701 Watertown Plank Road, Milwaukee, WI 53226; e-mail: ⟨dbragg@mcw.edu⟩.

References

1. Bland CJ, Holloway RL. A crisis of mission: faculty roles and rewards in an era of health care reform. Change. 1995;30–5.
2. Farrell TA. Teaching and managed care: are they compatible in the 21st century? Arch Ophthalmol. 1997;115:251–2.
3. Skeff KM, Bowen JL, Irby DM. Protecting time for teaching in the ambulatory care setting. Acad Med. 1997;72:694–7.
4. Bolognia JL, Wintroub BU. The impact of managed care on graduate medical education and academic medical centers. Arch Dermatol. 1996;132:1078–84.
5. Nordgren R, Hantman JA. The effect of managed care on undergraduate medical education. JAMA. 1996;275:1053–8.
6. Xu G, Wolfson P, Robeson M, Rodgers JF, Veloski JJ, Brigham TP. Students' satisfaction and perceptions of attending physicians' and residents' teaching role. Am J Surg. 1998;176:46–8.
7. Xu G, Hojat M, Veloski JJ, Gonnella JS. The changing health care system: a research agenda for medical education. Eval Health Prof. 1999;22:152–68.
8. Schum TR, Yindra KJ, Koss R, Nelson DB. Students' and residents' ratings of teaching effectiveness in a department of pediatrics. Teach Learn Med. 1993;5:128–32.
9. Kaufman L, Rousseeuw PJ. Finding groups in data: an introduction to cluster analysis. New York: Wiley, 1990.
10. Stout W, Habing B, Douglas J, Kim HR, Zhang J. Conditional covariance-based multidimensionality assessment. Appl Psychol Meas. 1996;20:331–54.
11. Ullian JA, Bland CJ, Simpson DE. An alternative approach to defining the role of the clinical teacher. Acad Med. 1994;69:832–8.
12. Schum TR, Yindra KJ. Relationship between systematic feedback to faculty and ratings of clinical teaching. Acad Med. 1996;71:1100–2.
13. Palmer P. The Courage to Teach. San Francisco, CA: Jossey–Bass, 1997.

Six-year Documentation of the Association between Excellent Clinical Teaching and Improved Students' Examination Performances

CHARLES H. GRIFFITH III, JOHN C. GEORGESEN, and JOHN F. WILSON

With increasing fiscal pressures on academic medical centers, many institutions are moving towards mission-based financing, the notion that the clinical, research, and teaching missions must no longer depend upon cross-subsidization but must financially support themselves.[1] With this increased mission-specific accountability, there will be greater emphasis on measurable outcomes to justify the costs associated with the mission. In the realm of clinical teaching, the literature is replete with studies of qualities of excellent teachers,[2] studies of how to measure teaching,[3] and studies demonstrating that faculty development in teaching can influence clinical teachers' self-reported behaviors,[4] actual behaviors,[5] and teaching ratings.[6] However, for the most part, the fundamental outcome of teaching has been left unstudied: that is, does the quality of teaching actually influence student learning? Although this may seem a truism too obvious for investigation, despite the cherished belief of clinical teachers there is very little quantitative evidence that better teaching is associated with enhanced student learning.

We recently reported the first documentation of the association of students' learning with the relative teaching abilities of attending physicians[7,8] and residents.[9] In these studies of students and their clinical teachers over the academic years 1993–1995, we found that medical students who worked on their internal medicine or surgical clinical clerkships with our best clinical teachers scored significantly higher on post-clerkship examinations and even on the U.S. Medical Licensing Examination (USMLE) Step 2. Our findings have been replicated at the University of Michigan.[10] The only other study noting an association of teaching with learning, published in 1983, involved high school students in a remedial math class.[11] To our knowledge, this is the extent of the quantitative evidence in all the educational literature that better teaching is associated with better learning.

Our previous reports, however, had several limitations. For one, our measure of teaching "quality" was based only on students' ratings. One can argue (as we did in those articles) that the learners are the best judges of the learning climate. Even though we controlled in our analysis for prior student academic achievement (USMLE Step 1 scores), it was possible that students especially excited about internal medicine scored better on internal medicine examinations and, in their enthusiasm, rated their instructors higher, with a spurious association of examination performance and teaching rating. Second, though statistically significant, our effect sizes were modest, amounting to one-sixth to one-seventh of a standard deviation on a test, or, for example, three points on the USMLE Step 2. Third, these studies encompassed only two academic years, and a limited number of teachers and students. Because this sample was small, we included in the analysis all teachers regardless of the numbers of students they worked with, even those with few teaching ratings. Though we were gratified to demonstrate an association between teaching and learning, our results may have been attenuated by the small sample and the inclusion of in the analysis of all teachers, regardless of numbers of teaching evaluations (teachers with imprecise measures of their teaching ability).

Therefore, the purpose of this project was to refine the method of our previous studies by using a larger sample of students and attending physicians, more precise measures of teaching ability, and a way of disentangling the potential confounders of raters and teachers. Our formal hypothesis was that students who are exposed to our highest-rated attending physicians during their internal medicine clerkship will score better on end-of-clerkship examinations and on the USMLE Step 2.

Method

This work represents an extension of the data set from our previous reports,[8,9] extending the sample size from two academic years to six. The study design, a prospective cohort study, involves data on students and their attending physicians, and notes the association of the students' examination performances with the "quality" of the attending physicians to whom they were exposed. The participants were all third-year medical students at the University of Kentucky College of Medicine, over the academic years 1993–1999 and their attending physicians on the inpatient general medicine services.

To give the reader a sense of the structure of our clerkship, students in the third year spend eight weeks on general medicine inpatient services, four at our university hospital, and four at our affiliated Veteran's Affairs hospital. A team consists of an attending physician, a supervising junior or senior medicine resident, two first-year residents, and two students. Importantly, students, housestaff, and attendings are randomly and independently assigned to the services (we do not take requests by students for specific attendings). Attending physicians may be either general internists or specialists. Note that students are exposed to new and different attending physicians and housestaff in each of the two four-week components of the clerkship. Ambulatory medicine is part of a separate primary care clerkship and is not included in the study. Attending physicians usually participate in or observe daily management rounds, and have formal separate teaching rounds three times per week, ideally focused on one or two patients on the service, usually at the bedside.

Our model for how teaching might influence students' learning was not that students would be influenced by the average teaching ability of all the instructors they worked with, but rather that students' learning would be enhanced by individual outstanding instructors who, in the learning climate they engender and the inspiration they provide each day, stimulate students to be excited about clinical medicine, resulting in students' learning not only throughout the clerkship but throughout all their clinical rotations. Therefore, we explored the associations of students' learning with exposures to particularly outstanding (or poor) attending physicians, rather than with the average ability of their two attending physicians.

In our prior studies,[8,9] we simply defined "best" and "worst" attending physicians as those with the highest and lowest teaching evaluations, as rated by students. However, as mentioned in our introduction, this could lead to a confounding of teaching rating with examination performance by a student who may perform better (and rate the physician attending higher) because of interest in internal medicine. Therefore, for this study, we elected to pursue an alternative method of identifying teaching quality. We surveyed a consensus panel of third- and fourth-year residents at our institution who had also been medical students here. These individuals would have had five to six years of exposure to the clinical teachers at out university, working with a great majority of them. We also chose former students who were residents because they would be

TABLE 1. Least-square mean results on the NBME Subject Exam in Medicine and USMLE Step 2 for 484 Students Working with Internal Medicine Attending Physicians Rated on their Teaching as High, Neither High Nor Low, and Low, University of Kentucky College of Medicine, 1993–1999*

Attending Physicians' Ratings†	No. of Students Who Worked with an Attending of this Level	NBME Subject Exam in Medicine Score (R^2 = 0.44)		Total USMLE Step 2 Score (R^2 = 0.57)	
		Score (SD)	*p* (between Scores)	Score (SD)	*p* (between Scores)
High	219	491 (112)	.007	207 (23)	.015
Neither high not low	220	463 (112)	—	203 (22)	—
Low	45	464 (90)	—	199 (22)	—

*Least-square mean results, which represent the predicted score for a student on the test, are controlled for USMLE Step 1 score.

†High- and low-rated attending physicians were those so rated by consensus of a panel of 15 residents who had formerly been students at the University of Kentucky College of Medicine.

‡Eighteen students who had worked with both a high-rated and a low-rated attending were excluded.

most familiar with the special needs and expectations of our internal medicine clerkship. We gave these residents a list of all faculty in internal medicine who had supervised more than five medical students during the six-year period. The threshold of five students was chosen because this was the number of evaluations we calculated were needed to achieve conventional standards of reliability for our clerkship's teaching evaluation form (greater than 0.80), and it would help identify those attending physicians for whom we had precise measures of their teaching ability. We asked the residents to confidentially rate faculty "high" if they would expect them to be rated among our best teachers, "low" if they were among our worst teachers, and "medium" if they would be in between. A priori, we defined "best" attending physicians as those that were named high rated instructors by 80% of the residents (at least 12 of the 15 residents) and were not mentioned as a low-rated instructor by any resident. Conversely, realizing the tendency for learners to rate even the worst instructors at least mediocre, we defined the "worst" attending physicians as those who were rated in the low category by at least five of the 15 residents, and who were not rated high by any of the residents.

For this study, students' evaluations of attending physicians' teaching quality were also collected over the six years, as further evidence of the validity for our consensus panel opinion (one would expect the instructors who were highly rated by residents' consensus to also have high teaching ratings if the consensus process is valid). Our measure of attending physicians' teaching quality was from confidential, end-of-month student evaluations, which were completed prior to the students' receiving their grades. The form consists of 16 items on a five-point Likert-type scale (1 = strongly disagree, 5 = strongly agree). Items included ratings of teaching skills and ability, rapport with learners and patients, overall rating, and ratings of their role modeling. The coefficient alpha for the evaluation form is .96. This means that there is a high degree of internal consistency among items for rating teaching, and that the instrument is a reliable measure of teaching. However, this also means that inter-item correlations are very high, for our form .75 to .95, which is not unusual for measures of clinical teaching.[12] Because of the high inter-item correlations, we used the mean rating across all items as one measure of teacher "ability." The overall rating an instructor was assigned in our data set was the mean of all the ratings from the students he or she worked with in the six academic years.

Our analysis used multiple regression approaches from the general linear model.[13] Our dependent variables were scores on the National Board of Medical Examiner's (NBME) subject examination in medicine, taken at the end of the clerkship, and USMLE Step 2 scores. Independent variables included dummy coded variables for different categories of attending exposure (i.e., high-rated versus low-rated versus neither high- nor low-rated attending physician exposure). We also included USMLE Step 1 scores in the model as a control variable for prior student academic achievement.

Results

Data were collected from 502 third-year medical students (100% of students) over the six academic years. We excluded 18 students who had worked with both a high-rated and a low-rated instructor (as our model was less clear about how this interaction might influence student learning), for a final sample of 484. A total of 46 attending physicians had more than five student evaluations over the six-year period and were included in the list that the consensus panel rated.

Overall, ten faculty met the criteria to be rated "high." Eight of the ten were rated high by all residents, and the other two by 13 and 14 residents, respectively. Four of the ten were general internists. Eight were men and two were women, which reflects our faculty demographics. Five faculty met our consensus criteria for a "low" rating, including one general internist and one woman.

Teaching evaluations were received from 96% of the students. The overall mean teaching rating of the teachers rated high was 4.68 on the five-point scale (SD = 0.22, range 4.23–4.94). For the "medium" group, the mean teaching rating was 4.34 (SD = 0.32, range 3.4–4.92). For the "worst"-rated attending physicians, the mean rating was 3.56 (SD = 0.48, range 3.06–4.21). Mean differences between groups were highly statistically significant ($p < 0.001$). Forty-five students had had exposures to at least one low-rated and no high-rated attending physician; 219 had had exposures to at least one high-rated and no low-rated attending physician; and 220 had had exposures to neither a high- nor a low-rated attending physician. Our high-rated attending physicians were more often attending physicians on the general medicine inpatient services than were the low- or medium-rated faculty, hence the disparity in numbers of students per faculty.

Table 1 presents the least-square mean scores on the post-clerkship NBME subject examination in medicine and on USMLE Step 2, depending on exposure to high-, low-, or medium-rated instructors (least-square means are predicted means adjusted for USMLE Step 1 scores). As can be seen, students who worked with at least one of our consensus panel's highly rated instructors scored significantly higher on the post-clerkship NBME examination in medicine and USMLE Step 2.

Conclusions

Our findings once again confirm the association of better clinical teaching with better student examination performance, demonstrating in a quantitative fashion the outcomes of teaching. The effect sizes in this current project are much more substantial than those in our prior reports, amounting to one-fourth to one-third of a standard deviation, or, for example, up to seven or eight points on USMLE Step 2, versus the one-sixth to one-seventh standard deviation effect sizes of our prior reports. We attribute our stronger conclusions to the more refined method of this current project.

First, our previous reports included all instructors, regardless of the numbers of students they had taught, and therefore all faculty were eligible to be included in the high- or low-rated category even if they had few student evaluations. For example, we may have included in our high category those faculty with only two or three ratings that were all high, when over time their ratings might have regressed towards a more stable mean that did not qualify them as such. In essence, we were categorizing some instructors as better or worse by using imprecise measures of their teaching ability. This imprecision would tend to add background "noise" to the analysis, attenuating our findings and effect sizes. Second, we disentangled learner outcomes from ratings by learners with our residents' consensus panel. As shown, attending physicians who were rated highly by the residents' consensus panel had significantly higher teaching ratings than did the medium- and low-rated instructors. Our previous method, relying on categorization solely by teaching rating, may have led to the exclusion of some otherwise excellent clinical teachers simply because they did not quite meet the "top 20% of student evaluations" we had required in our previous report to be considered a highly rated instructor.

Our findings seem to indicate that clinical teaching has an influence on outcomes, such as performance on USMLE Step 2. One might wonder how a short four-week exposure in a single discipline could influence USMLE Step 2 scores to such a degree, given that USMLE Step 2 comprises a wide variety of disciplines. Our answer is suggested by our model. From our experience as learners, the influence of a single outstanding instructor on one's approach to learning should not be underestimated. We suspect that the best teachers do not necessarily impart more factual information (facts which may be obsolete in a few years), but rather they engender a learning climate that makes learning fun, enjoyable, and exciting. They may do this by their example, by modeling the process of lifelong learning, by the joy they bring to their teaching, or by combinations of qualities such as these. Regardless, the learner's approach to learning is in some fundamental way changed, carrying over to the other clerkships and, we hope, to residency and beyond. Further studies should investigate the influence of outstanding teachers on life-long learning.

Several limitations to our study should be kept in mind as one interprets our results. This is a single-institution, single-discipline study, and certainly national studies are needed to assert the generalizability of our findings, as well as studies in other disciplines. In addition, our study focused on but one outcome measure, students' performances on NBME-type examinations, which measure but one aspect of clinical ability (knowledge). Future research should investigate the influence of teaching on other student outcomes, such as clinical skills, attitudes towards patients and the profession, and doctor–patient communication and relationships. Finally, this project's method did not lend itself as well to measuring the influence of residents' teaching on students' outcomes, so further studies are needed.

Nevertheless, despite these limitations, we conclude that attending physicians' teaching quality can have a measurable impact on students' examination performances. We therefore believe it is possible to begin considering learners' outcomes as an important measure of faculty's teaching ability, perhaps (with more study) an important addition to teaching portfolios and promotion dossiers. But even more, we believe our findings add to the growing literature on the critical importance of the educational mission that indicates students' learning would be jeopardized if the educational mission were to be compromised for fiscal reasons.

This research was supported by a grant from the National Board of Medical Examiners Research Fund, 56-9798.

Correspondence: Charles H. Griffith III, MD, MSPH, Department of Internal Medicine, University of Kentucky College of Medicine, Kentucky Clinic K512, Lexington, KY 40536-0284.

References

1. Watson RT, Romrell LJ. Mission-based budgeting: removing a graveyard. Acad Med. 1999;74:627–40.
2. Irby DM. What clinical teachers in medicine need to know. Acad Med. 1994;69: 333–42.
3. Speer AJ, Elnicki DM. Assessing the quality of teaching. Am J Med. 1999;106: 381–4.
4. Skeff KM, Stratos GA, Bergen MR, Sampson K, Deutsch SL. Regional teaching improvement programs for community-based teachers. Am J Med. 1999;106:76–80.
5. Wilkinson L, Sarkan RT. Arrows in the quiver: evaluation of a workshop in ambulatory teaching. Acad Med. 1998;73(10 suppl):S67–S69.
6. Skeff KM, Stratos GA, Berman J, Bergen MR. Improved clinical teaching: evaluation of a national dissemination program. Arch Intern Med. 1992;152:1156–61.
7. Blue AV, Griffith CH, Wilson JF, Sloan DA, Schwartz RW. Surgical teaching quality makes a difference. Am J Surg. 1999;177:86–9.
8. Griffith CH, Wilson JF, Haist SA, Ramsbottom-Lucier M. Relationships of how well attending physicians teach to their students performance and residency choice. Acad Med. 1997;72(10 suppl):S118–S126.
9. Griffith CH, Wilson JF, Haist SA, Ramsbottom M. Do students who work with better housestaff in their medicine clerkship learn more? Acad Med. 1998;73(10 suppl):S57–S59.
10. Stern DT, Gill A, Gruppen LD, Grum CM, Woolliscroft JO. Do the students of good teachers learn more? Abstract presented at Research in Medical Education Annual Meeting, Washington, DC, November, 1997.
11. Friedman M, Stamper C. The effectiveness of a faculty development program: a process-product experimental study. Rev Higher Educ. 1983;7:49–65.
12. Marriott DJ, Litzelman DK. Student's global assessments of clinical teachers: a reliable and valid measure of teaching effectiveness. Acad Med. 1998;73:572–4.
13. Cohen J, Cohen P. Applied multiple regression correlation analysis for the behavioral sciences. Hillsdale, NJ: Lawrence Erlbaum Associates, 1983.