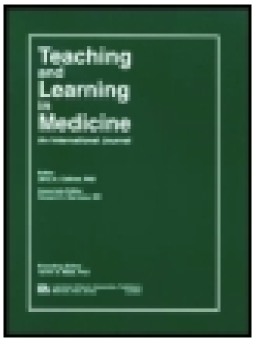

Teaching and Learning in Medicine

ISSN: 1040-1334 (Print) 1532-8015 (Online) Journal homepage: http://www.tandfonline.com/loi/htlm20

How Much Do Differences in Medical Schools Influence Student Performance? A Longitudinal Study Employing Hierarchical Linear Modeling

Kent Hecker & Claudio Violato

Teaching and Learning in Medicine, 20(2), 104–113

ISSN: 1040-1334 print / 1532-8015 online
DOI: 10.1080/10401330801991915

Routledge
Taylor & Francis Group

RESEARCH BASIC TO MEDICAL EDUCATION

How Much Do Differences in Medical Schools Influence Student Performance? A Longitudinal Study Employing Hierarchical Linear Modeling

Kent Hecker and Claudio Violato
Medical Education Research Unit, University of Calgary, Calgary, Alberta, Canada

***Background*: Medical school curricula have undergone considerable change in the past half century.There is little evidence, however, for the impact of various curricula and educational policies on student learning once incoming performance and the nonrandom nature of students nested within schools has been accounted for. *Purpose*: To investigate effects of school variables on United States Medical Licensing Examination (USMLE) Step 1–3 scores over an 11-year period (1994–2004). *Methods*: Using Association of American Medical Colleges and USMLE longitudinal data for 116 medical schools, hierarchical linear modeling was used to study the effects of school variables on Step 1–3. *Results*: Mean unadjusted between school variance was 14.74%, 10.50%, and 11.25%, for USMLE Step 1–3. When student covariates were included, between-school variation was less than 5%. The proportion of variance accounted for in-student-level performance by the covariates ranged from 27.58% to 36.51% for Step 1, 16.37% to 24.48% for Step 2, and 19.22% to 25.32% for Step 3. The variance accounted for in student performance by the student covariates ranged between 81.22% and 88.26% for Step 1, 48.44% and 79.77% for Step 2, and 68.41% and 80.78% for Step 3. School-level variables did not consistently predict for adjusted mean school Step performance. *Conclusions*: Individual student differences account for most of the variation in USMLE performance with small contributions from between-school variation and even smaller contribution from curriculum and educational policies.**

Calls for reform have dominated medical education for the past quarter century.[1–6] This is in response to a complex set of factors including advances in cognitive learning theory and application to medical education, the increase of biomedical sciences information, economic changes in the cost and delivery of health care, and social and cultural pressures due to the realization that the biomedical paradigm emphasizing curative medicine of the 1960s and 1970s is not attainable. In 2006, the *New England Journal of Medicine* dedicated journal space to issues in medical education to "support the fundamental restructuring of medical education needed today,"[3] (p. 1343) and the editor of *Academic Medicine* has suggested the need for a modern Flexner report to address the state of medical education in the United States and Canada.[7]

A plethora of educational changes have been suggested; the majority of empirical studies of "innovations" have reported mainly small effect sizes and nonsignificant differences in outcomes, however.[8–14] The few studies that have systematically evaluated the impact of varying curricula have failed to show substantive differences between curricula in outcomes.[15]

Curriculum, defined as "all the learning which is planned and guided by the school, whether it is carried on in groups or individually, inside or outside the school"[16] (p. 16), is the most susceptible to reform and change perhaps because it is the most recognizable and intuitive element of the educational process (other elements include teachers, students, and physical facilities such as laboratories, libraries, etc.). In United States and Canada, the development and implementation of curricula from kindergarten to professional programs have been heavily influenced by the constructivist (i.e., learning and knowledge is built on prior knowledge and cognitive structures), student-centered educational theories of John Dewey. Moreover, recent psychological theory has focused on constructivist and active (active participation of the learner) approaches to learning. How to apply these approaches to classroom practice and curriculum structure to improve learning, however, is still uncertain. Many of these recent innovations have been based on educational theories that have lacked empirical evidence[17] or indeed may have even ignored evidence confuting the theories (e.g., problem based, experiential, and inquiry based learning).[18]

American and Canadian medical education has been classified into five major curricular approaches implemented since the founding of the first medical school, the College of Philadelphia, in 1765.[19] These are the apprenticeship model (1765–present), discipline-based model (1871–present), organ-system-based model (1951–present), problem-based learning (PBL) model (1971–present), and clinical presentation (CP)-based model (1991–present).

Received 20 December 2007.
We gratefully acknowledge the AAMC and NBME for support of this project. Both authors contributed to the entire article.
Correspondence may be sent to Kent Hecker, 3330 Hospital Drive NW, Calgary Alberta, T2N 4N1, Canada. E-mail: kghecker@ucalgary.ca

The discipline-based and organ-system-based models are currently the most prevalent approaches. These can be traced back to Flexner's report on medical education, which stressed the scientific nature of medicine, with emphasis on basic sciences education and hypothesis-driven reasoning, combined with clinical experience.[20] The prototypical 4-year curriculum was first proposed by the Council on Medical Education and was strongly encouraged by Flexner.[20]

PBL and CP curricula grew out of the application of cognitive sciences to learning and the clinical reasoning theories and research to guide the development of students from medical novice to medical expert. In its ideal form, PBL is

> a curriculum of carefully selected and designed problems that demand from the learner acquisition of critical knowledge, problem-solving proficiency, self-directed learning strategies, and team-participation skills. Students work in small groups, generate hypotheses about the case and learning objectives, work outside of class to fill these deficiencies, then reconvene to teach each other and solve the problem.[21] (p. 300).

The justification for the use of PBL in medical education, specifically in the preclinical years,[22] is that it (a) provides exposure to clinical reasoning, (b) provides an organization of knowledge that has been theorized to aid students in the application of basic science knowledge to clinical problems,[12] (c) enhances transfer of content to new problems,[23,24] and (d) provides an environment that encourages self-directed learning and team working skills deemed to be important in medical practice.

The clinical presentation curriculum arose in the early 1990s as a result of the belief that how medical knowledge is taught in traditional curricula—organized by discipline, systems, or diseases and subsequently stored—is not how medical experts think or transmit knowledge from memory to solve clinical problems.[25,26] A clinical presentation is defined as "the ways in which a person, group of people, community, or population present to a physician" (p. 188) and the underlying assumptions is that the way the human body reacts to an infinite number of insults is finite and stable.[25] To date, there have been 125 ± 5 clinical presentations identified.[27]

Among educational theorists and researchers, assessing the between and within school effect on student performance is called school effectiveness research or educational effectiveness research that encompasses three broad areas: (a) school effects studies, which analyze the effects of schools on student and school performance through input–output studies and/or hierarchical linear model (HLM) or multilevel models; (b) effective schools studies, which attempt to identify school-level processes related to effective schooling through multimethod research designs studying classrooms and school-level effects as well as school case studies; and (c) school improvement studies, which can be prospective in nature, focusing on models that can affect school change.[28]

The ideal medical school effectiveness study should include (a) a large number of schools to encompass the natural variance in school characteristics, (b) large samples of students, (c) standardized assessment of student performance before entry into medical school (i.e., Medical College Admissions Test [MCAT]) to adjust for intake differences between schools, (d) standardized assessments measuring outcomes (i.e., United States Medical Licensing Exam [USMLE] Steps 1, 2, 3), (e) longitudinal data, (f) school-level variables (curriculum type, policies), and (g) the appropriate data analyses techniques, such as multilevel modeling, that recognizes the hierarchical nature of the data.[28,29]

Comprehensive data for medical school effectiveness research have been collected by the Association of American Medical Colleges (AAMC) and the National Board of Medical Examiners (NBME). These data include (a) student demographics (age, sex, underrepresented minority status [URM]) and performance entry data (MCAT scores, undergraduate grade point average [UGPA]); (b) medical student performance in the form of the USMLE Step 1 (typically taken after the 2nd year of medical school), Step 2 (typically taken during the 4th year of medical school), and Step 3 (typically taken during the latter part of the 1st or early part of the 2nd postgraduate year); and (c) school variables including curriculum structure, educational policies such as use of Step scores for promotion and graduation, and educational activities. With such a database, the between- and within-school variation on the outcome Step scores, as well as the between-year variation, can be modeled and the impact of school-level variables can be assessed. The study presented here employed such a comprehensive database within an HLM approach.

Medical school curriculum (course reorganization and pedagogical techniques) has undergone considerable change in the past half century, with the majority of schools claiming continual curriculum reform.[30] There is little evidence, however, for the impact of various curricula on student learning once incoming performance—prior achievement and MCAT scores—and the nonrandom nature of students nested within schools have been accounted for. Assessment of medical school curricula has been primarily cross-sectional and within one school evaluating various curricula. The main purpose of our study, then, was to assess the effects of school-level variables, specifically curriculum and educational policies, on preclinical and clinical measures of student performance.

Hierarchical linear modeling allows us to identify within- and between-school variations on performance while accounting for demographic and prior performance variables. School-level variables can then be modeled to determine how much varying curricula or educational policies can account for between school differences.

METHOD

Student data were initially obtained by Collin from the AAMC and the NBME and used in her dissertation, which assessed the predictive validity of the MCAT and the relationships between aptitude for medicine, academic achievement and performance in medicine.[31] This extensive data set contained

TABLE 1
School Level Independent Variables and Definitions

Abbreviation	School Variable
SPUBPRI	Public or private institution
SMEANGPA	School's mean undergraduate GPA for incoming medical students
SBS	School's mean biological sciences score on the MCAT subtest
SPS	School's mean physical sciences score on the MCAT subtest
SVR	School's mean verbal reasoning score on the MCAT subtest
SSYSTEMS	Dummy-coded curriculum variable—comparing systems based to discipline based
SMIXEDCU	Dummy-coded curriculum variable—comparing discipline-based 1st year and systems-based 2nd year to discipline based
SOTHERCU	Dummy-coded curriculum variable—comparing multitrack or other curricula to discipline based
SPBL	Dummy-coded curriculum variable—comparing problem-based learning curricula to discipline based
SNEREGIO	Dummy-coded region variable—comparing Northeastern schools to Southern schools
SCENREGI	Dummy-coded region variable—comparing Central schools to Southern schools
SWESTREG	Dummy-coded region variable—comparing Western schools to Southern schools
SS1RECOR	Dummy-coded Step 1 usage variable—comparing must only record a score to passing score required for promotion
SS1PASSS	Dummy-coded Step 1 usage variable—comparing passing score required for graduation to passing score required for promotion
SS1SCORE	Dummy-coded Step 1 usage variable—comparing must record a passing score on each section to passing score required for promotion
SS1OPTIO	Dummy-coded Step 1 usage variable—comparing taking the exam is optional to passing score required for promotion
SS2PASSF	Dummy-coded Step 2 usage variable—must record a passing score on each section to passing score required for graduation
SS2RECOR	Dummy-coded Step 2 usage variable—must only record a score to passing score required for graduation
SS2SCORE	Dummy-coded Step 2 usage variable—comparing must record a passing score on each section to passing score required for graduation
SS2OPTIO	Dummy-coded Step 2 usage variable—comparing taking the exam is optional to passing score required for graduation
SEDACTIV	Composite variable for educational innovations listed in AAMC Curriculum Directories

Note: MCAT = Medical College Admissions Test; AAMC = Association of American Medical Colleges.

anonymous information for students who applied for medical school from 1991 to 2001 ($N = 859{,}710$) including UGPA; MCAT subtests—Biological Sciences (BS), Physical Sciences (PS), Verbal Reasoning (VR) and Writing Sample (WS); USMLE Step 1, 2, and 3; sex; age; URM; and school attended, for all accredited medical schools registered with AAMC. Data were included if there were recorded matriculant scores for the MCAT subtests, UGPA, and recorded first-time scores on USMLE Step 1, Step 2, and Step 3. The data are organized by year of entry into medical school. The final data set contained longitudinal data for 8 years (1992–1999) for Step 1 ($n = 104{,}983$) and 2 ($n = 101{,}879$), and 7 years (1992–1998) for Step 3 ($n = 77{,}283$). This corresponded to the first cohort (1992) taking Step 1 in 1994 and the final cohort (1998) taking Step 3 in 2004.

School curricular information was obtained from the *AAMC Curriculum Directories*[32–39] and the 2000 *Academic Medicine* September supplement.[30] For educational research using HLM, sufficient sample sizes for the school level should be more than 50 schools and more than 30 students/school for unbiased estimates of first-level regression coefficients and variances and their standard errors, and second-level standard errors.[40,41] Therefore, as these values were used as a guideline for inclusion, some schools dropped below 30 students/year. For Step 2 as an outcome, 1 school dropped below 30 students (28) for 1 year; for Step 3, there were 4 years where schools dropped below 30 students/year. Three of the years had only 1 school drop below 30 students/year, and the final year (1998), 13 schools dropped below 30 students/year (the lowest was 11 students/year). These data were included, however, as an increase in groups permits the student-level number to decrease while still achieving unbiased estimates of the regression coefficients and variances and their standard errors. Schools with several campuses were either combined or kept separate based on the AAMC/NBME data classification. There were 116 schools included in the final analyses.

TABLE 2
Number and percentage of medical schools classified by curricular structure by year

Year	Discipline Based (%)	Organ Systems Based (%)	Discipline Based 1st Organ Systems 2nd (%)	Other (%)	Problem-Based Learning (%)
1992	76 (65.5)	6 (5.2)	14 (12.1)	8 (6.9)	12 (10.3)
1993	71 (61.2)	7 (6.0)	16 (13.8)	9 (7.8)	13 (11.2)
1994	69 (59.5)	7 (6.0)	16 (13.8)	10 (8.6)	14 (12.1)
1995	59 (50.9)	8 (6.9)	15 (12.9)	12 (10.3)	22 (19.0)
1996	55 (47.4)	9 (7.8)	15 (12.9)	12 (10.3)	25 (21.6)
1997	50 (43.1)	10 (8.6)	17 (14.7)	12 (10.3)	27 (23.3)
1998	48 (41.4)	11 (9.5)	19 (16.4)	10 (8.6)	28 (24.1)
1999	47 (40.5)	12 (10.3)	20 (17.2)	8 (6.9)	29 (25.0)

The *AAMC Curriculum Directory*[32–39] contains institutional grading evaluation characteristics (USMLE Step 1 requirements, USMLE Step 2 requirements), instructional innovations (use of self instruction, use of computer-assisted instruction, ambulatory primary care clerkships, use of standardized patients, peer review or clinical practice available), public versus private status, AAMC region, and school curricula (required courses, clerkships and educational intervention) for all accredited medical schools.

The *Academic Medicine* September supplement provided educational and curricular information for 118 of the 125 in the United States. Using this information and building on the curricular types identified by the *AAMC Curriculum Directory*;[32–39] Ripkey, Swanson, and Case;[15] and Papa and Harasym,[19] five curricular approaches were identified and coded:

1. Disciplines based: Courses such as anatomy, physiology, biochemistry, and genetics are taught in the 1st year and pathology, neurosciences, and pharmacology are present in the 2nd year.
2. Organ-system based: The disciplines are taught in the respective organ system, and courses such as renal, digestive, and endocrine are evident in the first 2 years.
3. Discipline based in 1st year and organ system based in 2nd year: First-year courses typically consist of biochemistry, anatomy, physiology, and genetics, and 2nd-year courses consist of endocrine, renal, digestive, and so on.
4. Other/multitrack: Universities offering multitrack programs, listing courses such as Doctoring I and II without a definition of the content, and so on.
5. PBL: Programs that have an identified PBL component that has been made explicit in the curriculum directory. This might include courses such as "problem-based learning" or stated problem-based components in all the courses presented in the curriculum. For verification, the numbers of hours in tutorials or cases were referred to for clarification as PBL is primarily disseminated through the use of tutorials.

For the school-level variables collected, a dummy variable coding scheme was created for USMLE Step 1 requirements for advancement, USMLE Step 2 requirements for advancement, curricular organization, public/private status, and AAMC region (Table 1). The refferant variable coded to zero was the subcategory with the largest number of units. Therefore, for k variables, there were k–1 dummy variables. For instance, for curricular type the referrant was discipline-based curricula, and each curricular structure was compared to this one. Schools reported either *yes* or *no* for five categories for instructional innovation over the 1992 to 1999 period. A composite variable was created for each school for each year where scores could range from 0 to 5, with a higher value representing a school reporting more of the instructional innovations. Other school-level variables included were mean school UGPA, MCAT-BS, MCAT-PS, and MCAT-VR scores.

HLM

To separate the effects on outcomes measured with student intake from those that were associated with the schools when estimating the impact of school, a two-level HLM analysis was employed with students nested within schools in three models with Steps 1, 2, and 3 of the USMLE as outcomes. In our analysis, a two-level hierarchical model was conducted because we were interested in identifying the within-year school-level variables that account for performance differences across multiple years.

The first model, one-way analysis of variance (ANOVA) with random effects, estimated by restricted maximum likelihood, provided estimates of the variance components at the student level and at the school level. These two values provided a measure of the relative size of the between-school variance.

To compare schools, the necessary adjustments for demographics and intake achievement must be made. Random coefficient models were tested secondly to identify the combination of sex, age of matriculation, URM, UGPA, and MCAT subscores to be included as covariates for within-school control on differences in school-level characteristics. These covariates were also assessed for their inclusion as either fixed or randomly varying across schools using a stepwise iterative process described in Moulder, Algina, and Julian.[42] Grand mean centering for each student-level covariate was used to predict for the adjusted school mean. Variance components calculated from the random coefficient model were used to calculate the variance accounted for by the covariates at the student level and school level.

The third approach, an intercepts as outcomes model, assessed the amount of variance accounted for by curriculum and educational policies in performance. A forward elimination procedure was used to identify significant school-level variables (those significant at $p < .05$ were retained). The percentage of total variance attributable to differences between schools was

TABLE 3
Variance components for students-nested-within-schools HLM analyses, proportion of school and student-level variance accounted for by student variables, and school-level variance accounted for by school variables

	Random Effects ANOVA			Random Coefficients Model			Intercepts as Outcomes Model		
	Step 1	Step 2	Step 3	Step 1	Step 2	Step 3	Step 1	Step 2	Step 3
1992									
School	77.47	66.04	35.98	10.02	16.49	6.92	8.26	14.75	4.52
Student	346.75	417.27	260.60	239.14	330.46	204.22	239.18	330.48	204.30
ρ(%)	**18.26**	**13.66**	**12.13**	**4.02**	**4.75**	**3.28**	**3.34**	**4.27**	**2.16**
1993									
School	74.30	65.42	43.41	12.67	18.00	9.33	10.60	17.68	8.02
Student	341.84	432.47	289.26	226.12	328.57	221.26	226.09	329.22	221.15
ρ(%)	**17.85**	**13.14**	**13.05**	**5.31**	**5.19**	**4.05**	**4.48**	**5.10**	**3.50**
1994									
School	72.58	62.72	38.58	11.87	17.01	8.48	11.03	16.77	6.32
Student	402.22	477.90	291.47	255.37	362.90	217.68	255.42	362.90	217.66
ρ(%)	**15.29**	**11.60**	**11.69**	**4.44**	**4.48**	**3.75**	**4.14**	**4.42**	**2.82**
1995									
School	65.64	67.59	39.36	8.80	13.68	11.06	8.80	12.28	7.62
Student	388.59	485.20	294.69	254.01	366.43	227.49	254.01	366.47	227.47
ρ(%)	**14.45**	**12.23**	**11.78**	**3.35**	**3.60**	**4.64**	**3.35**	**3.24**	**3.24**
1996									
School	68.29	57.89	32.80	10.85	14.58	9.55	10.13	13.08	7.26
Student	367.37	543.50	280.63	237.80	425.10	220.67	237.87	425.06	220.57
ρ(%)	**15.68**	**9.63**	**10.46**	**4.36**	**3.32**	**4.15**	**4.08**	**2.99**	**3.19**
1997									
School	62.23	47.99	30.17	7.30	15.27	8.43	6.06	13.91	6.53
Student	422.49	498.43	262.96	308.37	404.78	209.89	308.36	404.77	209.98
ρ(%)	**12.84**	**8.78**	**10.29**	**2.31**	**3.64**	**3.86**	**1.93**	**3.32**	**3.02**
1998									
School	65.09	37.78	26.13	12.22	18.11	8.25	11.54	14.96	5.59
Student	441.74	467.81	252.73	316.65	381.38	204.16	316.67	381.48	204.18
ρ(%)	**12.84**	**7.47**	**9.37**	**3.72**	**4.53**	**3.88**	**3.52**	**3.77**	**2.66**
1999									
School	59.88	38.69		10.70	19.95		10.58	15.46	
Student	498.23	477.98		360.82	399.75		360.78	399.70	
ρ(%)	**10.73**	**7.49**		**2.88**	**4.75**		**2.85**	**3.72**	

Note: ρ (the percentage of total variance attributable to differences between schools) = school variance/ (school variance + student variance) $\times$ 100. HLM = hierarchical linear model; ANOVA = analysis of variance.

determined by (school variance/ (school variance + student variance)) $\times$ 100.

Our study received approval from the Conjoint Health Research Ethics Board of the University of Calgary.

RESULTS

Medical School Curricula and Educational Policies

From 1992 to 1999, the most common curricular structure was discipline based. This curricular structure was also subject to the greatest change over the period, from 76 schools reporting this structure in 1992 to 47 in 1999 (see Table 2). The presence of a PBL curricula or stated curricular components increased the most over this period, from 12 schools reporting this structure in 1992 to 29 schools in 1999. Organ-systems-based curricula also increased over this period from 5% of the schools reporting this curricular structure in 1992 to 10% of the schools in 1999. Discipline-based 1st year and organ-systems-based 2nd year and multitrack curricula also fluctuated over the same period with no discernable pattern.

TABLE 4
Fixed effects results from intercepts as outcomes model for students-nested-within-schools HLM analyses

USMLE	
Step 1[a]	
1992	$\beta_{0j} = 207.14 - 2.35\text{SWEST} - 1.57\text{SS1PASSS} - 2.60\text{SS1OPTIO} + u_{0j}$
1993	$\beta_{0j} = 209.05 + 2.18\ \text{SPUBPRI} - 3.19\text{SS1OPTIO} + u_{0j}$
1994	$\beta_{0j} = 211.26 + 1.78\ \text{SPUBPRI} - 2.12\ \text{S1OPTIO} + u_{0j}$
1995	$\beta_{0j} = 213.27 + u_{0j}$
1996[b]	$\beta_{0j} = 216.98 + 2.22\ \text{SOTHERCU} + \mathbf{2.01\ SPBL} + u_{0j}$
1997[b]	$\beta_{0j} = 217.19 + 1.32\ \text{SPUBPRI} - \mathbf{1.87\ SSYSTEMS} - 2.54\ \text{SS1PASSS} + u_{0j}$
1998[b]	$\beta_{0j} = 216.84 + \mathbf{2.14\ SPBL} + u_{0j}$
1999	$\beta_{0j} = 216.79 + 1.82\ \text{SS1OPTIO} + u_{0j}$
Step 2[c]	
1992	$\beta_{0j} = 204.36 + 9.10\ \text{SMEANGPA} - 2.93\ \text{SWEST} + u_{0j}$
1993	$\beta_{0j} = 207.93 + 9.91\ \text{SS1SCCOR} + u_{0j}$
1994[b]	$\beta_{0j} = 209.81 + \mathbf{2.26\ SOTHERCU} + u_{0j}$
1995[b]	$\beta_{0j} = 211.09 - 1.67\ \text{SPUBPR} + \mathbf{2.06\ SPBL} + u_{0j}$
1996[b]	$\beta_{0j} = 212.85 - 1.84\ \text{SPUBPR} + \mathbf{1.81\ SPBL} - 2.45\ \text{SWEST} + u_{0j}$
1997[b]	$\beta_{0j} = 215.60 + \mathbf{2.27\ SPBL} - 2.70\ \text{SWEST} + u_{0j}$
1998	$\beta_{0j} = 217.14 - 1.90\ \text{SBS} - 2.11\ \text{SS2RECOR} + u_{0j}$
1999	$\beta_{0j} = 217.22 - 1.01\ \text{SPS} - 2.82\ \text{SNE} - 3.29\ \text{SWEST} - 3.11\ \text{SS2RECOR} + u_{0j}$
Step 3[d]	
1992	$\beta_{0j} = 207.25 - 2.53\ \text{SPUBPRI} - 0.97\ \text{SVR} + 1.42\ \text{SCEN} + u_{0j}$
1993[b]	$\beta_{0j} = 209.12 - 1.73\ \text{SPUBPRI} + \mathbf{2.43\ SSYSTEMS} + 2.24\ \text{SCEN} + u_{0j}$
1994	$\beta_{0j} = 208.81 - 1.97\ \text{SPUBPRI} + 2.32\ \text{SCEN} + u_0$
1995	$\beta_{0j} = 210.45 - 2.76\ \text{SPUBPRI} + 4.67\ \text{SMEANGPA} + 1.54\ \text{SCEN} + 0.95\ \text{SEDACTIV} + u_{0j}$
1996	$\beta_{0j} = 211.57 - 2.84\ \text{SPUBPRI} + 1.34\ \text{SCEN} + u_{0j}$
1997[b]	$\beta_{0j} = 212.74 - 1.73\ \text{SPUBPRI} + \mathbf{1.59\ SMIXEDCU} + 1.63\ \text{SCEN} + u_{0j}$
1998[b]	$\beta_{0j} = 213.08 - 2.48\ \text{SPUBPRI} + \mathbf{1.73\ SMIXEDCU} + 2.22\ \text{SCEN} + u_{0j}$

Note: AAMC = Association of American Medical Colleges; SCEN = AAMC central region; SWEST = AAMC west region; SNE = AAMC northeast region; SS1PASS = Step 1 passing score required for graduation; SS1OPTIO = taking the Step 1exam is optional; SS1SCCOR = must record a score for Step 1; SPUBPRI = Public or private school; SPBL = PBL curriculum; SMIXEDCU = discipline-based 1st-year systems-based 2nd-year curriculum; SOTHERCU = other or multitrack curriculum; SSYSTEMS = systems-based curriculum SMEANGPA = school mean GPA; SBS = school mean Biological Sciences score; SPS = school mean Physical Sciences score; SVR = school mean Verbal Reasoning score; SS2RECOR = must only record a score on Step 2; SEDACTIV = number of instructional innovations record by school.

[a] $n = 104{,}983$.

[b] Year when curriculum differences was a significant variable.

[c] $n = 101{,}879$.

[d] $n = 77{,}283$.

Schools also changed policies with respect to use of Step 1 and Step 2 requirements over this period. Specifically, more schools in 1999 used Step 1 scores for promotion and graduation than in 1992, and more schools required Step 2 scores for graduation in 1999 than 1992. Comparing Step 1–3 means for schools requiring Step 1 for promotion and graduation to those that did not, mean values were consistently within five points across all three Steps. Multivariate analysis of variance (dependent variables mean USMLE Step 1, 2, and 3 scores) indicated that, for Step 1 as an outcome, from 1992 to 1994 there were no significant differences between schools ($p > .05$), whereas from 1995 to 1998 there were differences with those schools not requiring Step 1 for promotion and graduation having higher Step 1 means ($p < .05$). Effect size calculations (d) demonstrated that the differences were small ($d = 0.11$– 0.19). There were significant differences between the two groups of schools for Step 2 and 3 scores, where schools who did not require Step 1 for promotion or graduation having higher Step 2 and Step 3 means ($p < .05$). Again, effect sizes were small ($d = 0.08$–0.20). Comparing Step 1 to 3 means for school policies on Step 2 requirements did not give rise to any discernable pattern; typically, the results were nonsignificant on all the Step scores. For those years where there were significant differences effect sizes were small ($d \leq 0.08$).

Random Effects ANOVA

The results of the ANOVA are summarized in Table 3 (random effects ANOVA). The between-school proportion of variance range for Step 1 scores was 10.73% to 18.26%, 7.47% to 13.66% for Step 2 scores, and 9.37% to 13.05% for Step 3 scores. Mean unadjusted between school variance was 14.74%, 10.50%, and 11.25% for USMLE Step 1–3 respectively, and these between-school proportions are consistent with Ripkey et al.[15] and Moulder et al.[42] From 1992 to 1999, for all three Step scores the overall trend was a decrease in the variances attributed to school differences with some minor fluctuations.

Random Coefficients Models

The student level covariates included were the MCAT subtests—BS, PS, and VR; cumulative UGPA; sex; age; and URM. The WS subtest was not included as a student-level covariate because it did not account for a large amount of variance in the analyses. This is consistent with previous predictive validity studies.[43,44] The variance accounted for in student performance by the student covariates ranged from 27.58% to 36.51% for Step 1, 16.37% to 24.48% for Step 2, and 19.22% to 25.32% for Step 3 (Figure 1). The proportion of the between-school variance that was accounted for by the student covariates ranged from 81.22% to 88.26% for Step 1, 48.44% to 79.77%[1] for Step 2, and from 68.41% to 80.78% for Step 3 (Figure 1). The between-school proportion of variance when student variables were employed as covariates decreased compared to the random effects ANOVA and varied between 2.88% and 5.31% for Step 1, between 3.32% and 5.19% for Step 2, and between 3.28% and 4.64% for Step 3 (Table 3, Random Coefficients Model columns). The MCAT variables contributed the largest proportion of variance in student performance (12.87–30.87%) for all analyses. Although other student variables (sex, URM, age at matriculation) were statistically significant ($p < .05$), their effect was relatively small (3.85% to 6.82%). The results summarized in Table 3 and Figure 1 show that a large proportion of the variance in student performance on the USMLE is attributable to student characteristics (mostly MCAT scores) and very little to school differences.

[1]There appears to be a trend of decreasing variance in Step 2 accounted for by student variables. However, the between-school variance calculated in the random effects ANOVA decreased over the same period from 13.66% to 7.49%, whereas the between-school variance calculated in the random coefficients model was relatively constant (3.32%–5.19%).

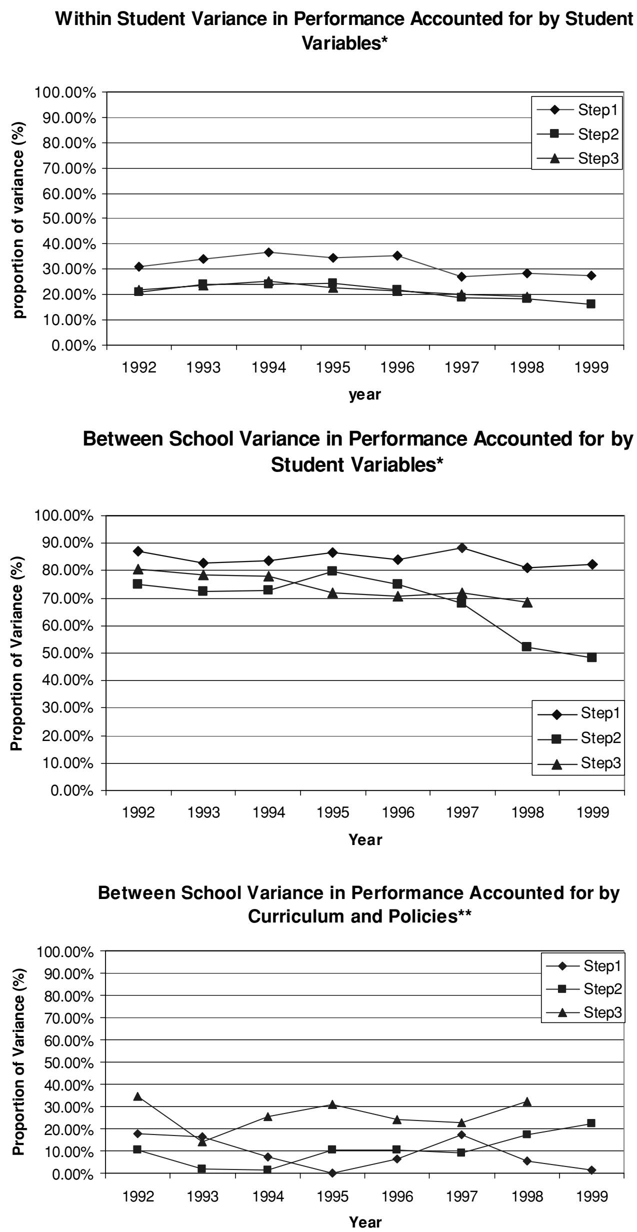

FIG. 1. Comparison of proportion of variance accounted for by student variables for within-student variance, between-school variance, and between-school variance accounted for by school-level variables, curriculum and educational policies. *Student variables – Medical College Admissions Test, undergraduate grade point average, sex, age, and underrepresented minority status; **Curriculum and educational policies variables, defined in Table 2.

Intercepts as Outcomes Models

The results of the intercept model are summarized in Table 4. The equations in Table 4 represent the linear combination of significant ($p < .05$) independent variables for the three dependent variables (USMLE Steps 1–3) by year. An intercepts as outcomes equation models differences in the adjusted mean (intercept, β_{0j}) for students in school j. Significant coefficients for each of the school-level predictors can be read as in simple regression where the outcome dependent variable is assessed against unit increases in the predictor while holding the other predictors constant. If the dummy variables were significant, then the predictor was assessed against the referent while holding the other predictors constant. For Step 1 results, for only 3 years (1996–1998) did curriculum differences between schools contribute to the equation. The most consistent variables influencing Step 1 performance over the entire period (1992–1999) were whether a passing score was required for graduation and whether taking the Step 1 exam was optional. Geographic location and private/public school status appeared in the regression equations (Table 4). Similarly, there were no consistent patterns of curriculum effects and educational innovations over the 8 years for Step 2: Some curriculum differences emerged during the period 1994–1997. Geographic location, private/public school, school-level GPA, MCAT scores, and Step 1 and Step 2 policies appeared as significant independent variables.

Conversely, a consistent pattern emerged from the Step 3 intercepts from 1992 to 1998: Private/public school effects emerged over the entire period as the most important independent variable in the equations. Some curriculum and educational innovation effects emerged sporadically in 1993, 1995, and 1997–1998. Other independent variables emerged sporadically over the years.

In concordance with the foregoing analyses, the between-school variance has been reduced to less than 5% over the period over the three Step measures. The school-level variables accounted for between 0%[2] and 17.55% of the school variance in Step 1, between 1.37% and 22.47% of the school-level variance in Step 2, and between 14.06% and 34.67% in Step 3 (Table 3, Intercepts as Outcomes Model columns, and Figure 1).

CONCLUSION

To assess the impact of medical school curriculum and educational policies on student performances over an 8-year period, a two-level hierarchical linear model was utilized. We have five main findings:

1. The majority of the variation between schools in Step 1–3 exams can be accounted for by incoming students differences, mostly MCAT scores. Student differences accounted for greater than 85% of the total variation within schools.
2. The mean unadjusted proportion of between school variance was small (<15%) for all three Steps.
3. Between 16% and 36% of the student variance for Step 1–3 overall can be accounted for by student variables.
4. Curriculum differences and school-level educational policies and educational innovations contributed only sporadically in the regression equations over the 8-year period.
5. Two consistent school variables that did contribute significantly to between school differences were geographic location and private/public status.

Our results from the two-level models are in concordance with Moulder et al.[42] and Ripkey et al.,[15] but we added the comprehensive analysis of the within- and between-school variance longitudinally by year. Moreover, we were able to show curriculum effects between schools were sporadic and accounted for a small proportion of the variance in USMLE Step 1–3 performance when we adjusted for student variables (i.e., MCAT scores).

The student demographic and performance variables accounted for 31.77% in Step 1, 21.10 % in Step 2, and 22.00% in Step 3 for the within-student variance. These values are similar to those reported by Julian[43] and Donnon, Paolucci, and Violato[44] In their meta analysis, Donnon et al. reported that total MCAT scores accounted for 36.00% of the variance in Step 1, 14.44% in Step 2, and 18.49% in Step 3. The convergence of these findings lends confidence to our results.

Various curricula or educational policies within schools have little differential impact on basic sciences concepts (Step 1); the application of medical knowledge, skills, and clinical sciences in the provision of supervised patient care (Step 2); and the application of medical knowledge and understanding of biomedical and clinical sciences in the provision of unsupervised patient care (Step 3). When there were significant curricular effects within Step scores, the effect was not consistent year to year. For example, schools with a PBL curricula in 1996 and 1998 scored significantly higher than Discipline-based curricula on Step 1, whereas schools with a Systems-based curricula scored significantly lower than schools with a Disciplines-based curricula in 1997 (Table 4). Although multivariate analysis of variance results indicated that there were significant differences between unadjusted mean Step 1–3 performance between schools that require Step 1 for promotion and graduation to those schools that didn't, the effect sizes were very small, and this school-level variable was not a consistent predictor of between school performance in the HLM results. Of the between-school differences, school variables that accounted for some variance (8–26%) on Steps 1–3 were primarily due to geographical region and private/public school designation. Privately funded schools on average scored significantly higher on Step 1 exams. Conversely, private schools performed on average lower than the public schools over the same 7-year period on Step 3. This has been reported by Moulder et al.[32] Perhaps the differences are due to continued emphasis on basic sciences (as measured by Step 1) by private schools while the public schools have shifted emphasis to clinical skills and procedures (Step 3). The differences may also be due to school policies about taking Step 1. For Step 2, there was no consistent private/public variation in performance. There was again little impact of curriculum or educational policies accounted for adjusted Step 2 score. In any case, the overall amount of variance

[2]This value—0% variance—is likely an artifact of calculating effect size in HLM, although the true variance is likely very close to 0.[45]

accounted for by these school factors was small in all three Steps.

Although our study has several strengths (large database, longitudinal over 10 years, includes most medical schools, hierarchical modeling), there are some limitations. First, we had to rely on stated curricula and policy for our classification. What schools advertise on Web sites or state in documents about their curricula and what is implemented in practice is not always consistent. Second, the dependent variables are licensing exam scores, which may not fully assess important physician characteristics such as empathy, problem solving, clinical reasoning,[46] and so on, that are learning objectives of some medical schools. The arguments that certain curricula provide greater exposure and education for these physician characteristics are difficult to assess between schools because of the lack standardized between school assessment tools meant specifically for these areas. Finally, we were not able to include teacher variables (e.g., teacher effectiveness, formal education in teaching methods, contact hours, etc.) in our analyses. Future research may well investigate the impact of such variables.

Notwithstanding these limitations, the differences between schools such as differing curricula do not account for much variation of student performance on USMLE exams. Most of the variance is because of student entry differences as measured using currently available standardized performance measures such as MCAT scores. The obvious conclusion, therefore, is that changing curricula in medical education reform[1,3,5–7] is not likely to have much impact in improvement in student achievement. Future work ought to focus on student characteristics and teacher characteristics such as teaching competency.

REFERENCES

1. Carraccio C, Wolfsthal SD, Englander R, Ferentz K, Martin C. Shifting paradigms: from Flexner to competencies. *Academic Medicine* 2002;77(5):361–7.
2. Pauli HG. Education and training of the physician: Undergraduate medical education and health care. *Journal of Medical Education* 1973;48(12):51–5.
3. Cooke M, Irby DM, Sullivan W, Ludmerer KM. American medical education 100 years after the Flexner report. *New England Journal of Medicine* 2006;355(13):1339–44.
4. The Blue Ridge Academic Health Group. *Reforming medical education: urgent priority for the academic health center in the new century*. Atlanta, GA: Robert W. Woodruff Health Sciences Center, 2003.
5. The Commonwealth Fund. *Training tomorrow's doctors: The medical education mission of academic health centers*. New York: The Commonwealth Fund, 2002.
6. Association of American Medical Colleges. *Educating doctors to provide high quality medical care: A vision for medical education in the United States*. Washington, DC: Association of American Medical Colleges, 2004.
7. Whitcomb ME. Medical education reform: Is it time for a modern Flexner report? *Academic Medicine* 2007;82(1):1–2.
8. Goodman LJ, Brueschke EE, Bone RC, Rose WH, Williams EJ, Paul HA. An experiment in medical education: A critical analysis using traditional criteria. *Journal of the American Medical Association* 1991;265(18):2373–6.
9. Moore GT, Block SD, Style CB, Mitchell R. The influence of the New Pathway curriculum on Harvard medical students. *Academic Medicine* 1994;69(12):983–9.
10. Mennin S, Kalishman S, Friedman CP, Pathak D, Snyder J. A survey of graduates in practice from the University of New Mexico's conventional and community-oriented, problem based tracks. *Academic Medicine* 1996;71:1079–89.
11. Distlehorst LH, Robbs RS. A comparison of problem-based learning and standard curriculum students: Three years of retrospective data. *Teaching and Learning in Medicine* 1998;10(3):131–7.
12. Albanese MA, Mitchell S. Problem-based learning: A review of literature on its outcomes and implementation issues. *Academic Medicine* 1993;68(1):52–81.
13. Vernon DT, Blake RL. Does problem-based learning work? A meta-analysis of evaluative research. *Academic Medicine* 1993;68:550–63.
14. Colliver JA. Effectiveness of problem-based learning curricula: Research and theory. *Academic Medicine* 2000;75(3):259–66.
15. Ripkey DR, Swanson DB, Case SM. School-to-school differences in step 1 performance as a function of curriculum type and use of step 1 in promotion/graduation requirements. *Academic Medicine* 1998;73(10 Suppl.):S16–S18.
16. Kerr J. The problem with curriculum reform. In J Kerr (Ed.), *Changing the curriculum* (pp. 13–38). London: University of London Press, 1968.
17. Colliver JA. Educational theory and medical education practice: A cautionary note for medical school faculty. *Academic Medicine* 2002;77(12 Pt 1):1217–20.
18. Kirschner PA, Sweller J, Clark RE. Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist* 2006;41(2):75–86.
19. Papa FJ, Harasym PH. Medical curriculum reform in North America, 1765 to the present: A cognitive science perspective. *Academic Medicine* 1999;74(2):154–64.
20. Flexner A. *Medical education in the United States and Canada. A report to the Carnegie Foundation for the advancement of teaching*. Boston: Updyke 1910.
21. Kinkade S. A snapshot of the status of problem-based learning in U.S. medical schools, 2003–04. *Academic Medicine* 2005;80(3):300–1.
22. Whitfield CF, Mauger EA, Zwicker J, Lehman EB. Differences between students in problem-based and lecture-based curricula measured by clerkship performance ratings at the beginning of the third year. *Teaching and Learning in Medicine* 2002;14(4):211–7.
23. Norman GR, Schmidt HG. The psychological basis of problem-based learning: A review of the evidence. *Academic Medicine* 1992;67(9):557–65.
24. Donnon T, Violato C. Medical students' clinical reasoning skills as a function of basic science achievement and clinical competency measures: A structural equation model. *Academic Medicine* 2006;81(10):120–3.
25. Mandin H, Harasym P, Eagle C, Watanabe M. Developing a "clinical presentation" curriculum at the University of Calgary. *Academic Medicine* 1995;70:186–93.
26. Mandin H, Jones A, Woloschuk W, Harasym P. Helping students learn to think like experts when solving clinical problems. *Academic Medicine* 1997;72(3):173–9.
27. Mandin H, Dauphinee WD. Conceptual guidelines for developing and maintaining curriculum and examination objectives: The experience of the Medical Council of Canada. *Academic Medicine* 2000;75(10):1031–7.
28. Teddlie C, Reynolds D (Eds.). *The international handbook of school effectiveness research*. New York: Falmer, 2000.
29. Rivkin SG, Hanushek EA, Kain JF. Teachers, schools, and academic achievement. *Econometrica* 2005;73(2):417–58.

30. Anderson MB. A guide to the 130 reports in this snapshot supplement to *Academic Medicine*. *Academic Medicine* 2000;75(9):Sx–xiv.
31. Collin VT. Predictive validity study of the medical college admission test using multiple regression and latent variable path analysis (LVPA) (Dissertation, Calgary (AB), University of Calgary), 2005.
32. Association of American Medical Colleges. *Curriculum directory 1992–1993* (21st ed.). Washington, DC: Association of American Medical Colleges, 1993.
33. Association of American Medical Colleges. *Curriculum directory 1993–1994* (22nd ed.). Washington, DC: Association of American Medical Colleges, 1994.
34. Association of American Medical Colleges. *Curriculum directory 1994–1995* (23rd ed.). Washington, DC: Association of American Medical Colleges, 1995.
35. Association of American Medical Colleges. *Curriculum directory 1995–1996* (24th ed.). Washington, DC: Association of American Medical Colleges, 1996.
36. Association of American Medical Colleges. *Curriculum directory 1996–1997* (25th ed.). Washington, DC: Association of American Medical Colleges, 1997.
37. Association of American Medical Colleges. *Curriculum directory 1997–1998* (26th ed.). Washington, DC: Association of American Medical Colleges, 1998.
38. Association of American Medical Colleges. *Curriculum directory 1998–1999* (27th ed.). Washington, DC: Association of American Medical Colleges, 1999.
39. Association of American Medical Colleges. *Curriculum directory 1999–2000* (28th ed.). Washington, DC: Association of American Medical Colleges, 2000.
40. Kreft IGG. *Are multi-level techniques necessary? An overview, including simulation studies*. Los Angeles: California State University, 1996.
41. Hox J, Mass C. Sufficient sample sizes for multilevel modeling. *Method* 2005;1:85–91.
42. Moulder BC, Algina J, Julian E. *Impact of medical school characteristics on the predictability of USMLE scores* [MCAT Monograph], 2004.
43. Julian ER. Validity of the Medical College Admission Test for predicting medical school performance. *Academic Medicine* 2005;80(10):910–7.
44. Donnon T, Paolucci EOP, Violato C. The predictive validity of the MCAT for Medical school performance and medical board licensing examinations: A meta-analysis of the published research. *Academic Medicine* 2007;82(1):100–6.
45. Raudenbush SW, Bryk AS. *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage 2002.
46. Violato C, Donnon T. Does the Medical College Admission Test predict clinical reasoning skills? A longitudinal study employing the Medical Council of Canada Clinical Reasoning Examination. *Academic Medicine* 2005;80:S14–16.