Influence of curriculum type on student performance in the United States Medical Licensing Examination Step 1 and Step 2 exams: problem-based learning vs. lecture-based curriculum

Cam Enarson[1] *& Liza Cariaga-Lo*[2]

Introduction The results of the United States Medical Licensing Examination Step 1 and 2 examinations are reported for students enrolled in a problem-based and traditional lecture-based curricula over a seven-year period at a single institution. There were no statistically significant differences in mean scores on either examination over the seven year period as a whole. There were statistically significant main effects noted by cohort year and curricular track for both the Step 1 and 2 examinations. These results support the general, long-term effectiveness of problem-based learning with respect to basic and clinical science knowledge acquisition.

Context This paper reports the United States Medical Licensing Examination Step 1 and Step 2 results for students enrolled in a problem-based and traditional lecture-based learning curricula over the seven-year period (1992–98) in order to evaluate the adequacy of each curriculum in supporting students learning of the basic and clinical sciences.

Methods Six hundred and eighty-nine students who took the United States Medical Licensing Examination Step 1 and 540 students who took Step 2 for the first time over the seven-year period were included in the analyses. *T*-test analyses were utilized to compare students' Step 1 and Step 2 performance by curriculum groups.

Results United States Medical Licensing Examination Step 1 scores over the seven-year period were 214 for Traditional Curriculum students and 208 for Parallel Curriculum students (t-value $= 1{\cdot}32$, $P = 0{\cdot}21$). Mean Step 2 scores over the seven-year period were 208 for Traditional Curriculum students and 206 for Parallel Curriculum students (t-value $= 1{\cdot}08$, $P = 0{\cdot}30$). Statistically significant main effects were noted by cohort year and curricular track for both the Step 1 and Step 2 examinations.

Conclusion The totality of experience in both groups, although differing by curricular type, may be similar enough that the comparable scores are what should be expected. These results should be reassuring to curricular planners and faculty that problem-based learning can provide students with the knowledge needed for the subsequent phases of their medical education.

Keywords *Curriculum; educational measurement/ *methods; education, medical; *licensure; problem-based learning; United States.

Medical Education 2001;35:1050–1055

Introduction

Over the last decade, educational approaches that promote active, self-directed learning have been increasingly integrated into the curricula of many medical schools. However, the concern remains that these approaches may not provide students with sufficient biomedical science knowledge to prepare them for the next phases of their education. Among the outcome measures used to assess the acquisition of basic and clinical knowledge, the results of the licensing examinations administered by the National Board of Medical Examiners continue to be of critical value to the academic medicine community.

Previous research has examined the effect of curriculum type on the National Board of Medical Examiners Part I and Part II examination scores. Farquar *et al.* compared the National Board of Medical Examiners

[1]Wake Forest University School of Medicine, Winston-Salem, USA
[2]Yale University Graduate School, USA

Correspondence: Cam Enarson, Associate Dean for Medical Education, Wake Forest University School of Medicine, Winston-Salem, North Carolina 27157–1037, USA. Tel.: 00 1 336 716 5168; Fax: 00 1 336 716 5807; E-mail: cenarson@wfubmc.edu

Key learning points

PBL provides for effective learning of basic and clinical science information.

Pre-matriculation basic science preparation is an important factor with respect to peformance on standardized examinations.

Longitudinal study of curricular outcomes is needed to assess the impact of curriculum type on subsequent performance along the educational continuum.

Part I performance for a matched sample of students in the two curricular tracks at Michigan State University College of Human Medicine.[1] Forty students from each track were matched by Medical College Admission Test scores. No significant difference in total test scores was observed between students in the two tracks. The impact of curriculum type on student performance on the the National Board of Medical Examiners Part I and Part II examinations was studied longitudinally at Rush Medical College by Goodman *et al.*[2] Five hundred and ninety-seven students from the matriculating classes of 1984–88 were compared with respect to Part I scores and 367 students from the matriculating classes of 1984–86 were compared for Part II scores. Eighteen students per year were chosen to enter the alternative, problem-based track following admission. The analyses included 72 alternative track students for Part I and 36 for Part II. The mean scores were adjusted for undergraduate science, grade point average, medical college admission test scores, age, gender and other degrees earned. No significant difference in adjusted scores were found for either the Part I or Part II examinations. Mennin *et al.* extended the analysis by comparing student performance on Steps 1, 2 and 3 at the University of New Mexico.[3] For Part I, the scores of 508 conventional track and 167 Parallel Curriculum Course (PCC) students from the classes of 1983–92 were compared. Four hundred and forty-seven conventional track and 144 Parallel Curriculum Course students from the classes of 1983–91 were compared on Part II scores. The Parallel Curriculum students included students who were selected directly into the PCC and students who were randomized into the PCC after requesting admission into this track. Students in the Conventional Track scored significantly higher on the Part I exam compared to their PCC counterparts (504 vs. 456 $P = 0{\cdot}0001$). Students who had expressed a preference for the PCC, but had been randomized into the conventional track, scored higher on Part I (mean 521) than students randomized into the PCC (mean 455) or directly selected into the PCC (mean 463). There was no statistically significant difference in Part II scores, although the mean scores for Parallel Curriculum Course students were higher than their conventional track colleagues (469 vs. 460, $P = 0{\cdot}29$). Distlehorst and Robbs compared United States Medical Licensing Examination Step 1 and 2 scores for three graduating classes (1994–96) from Southern Illinois University School of Medicine.[4] Forty-seven students from the PBL track and 154 standard curriculum students were included in the analysis. No difference in Step 1 scores was found for students in the problem-based learning curriculum as compared to students in the standard curriculum. Students in the problem-based curriculum had higher Step 2 scores than their standard curriculum counterparts. This difference became non-significant when scores were controlled for rural location of matriculating students, age at matriculation and overall Medical College Admission Test score. Blake *et al.* compared USMLE Step 1 and 2 scores for students enrolled in the revised curriculum at the University of Missouri-Columbia School of Medicine with the scores of students enrolled prior to revision of the curriculum at the same school.[5] Students in the new curriculum demonstrated significant improvement on both the Step 1 and Step 2 exams as compared to students who had completed the school's traditional curriculum.

The National Board of Medical Examiners Part I and Part II examinations were replaced by the new United States Medical Licensing Examination Step 1 and Step 2 in 1992. This paper reports the United States Medical Licensing Examination Step 1 and Step 2 results for students enrolled in one of two curricular tracks, problem-based or traditional lecture-based, over the seven-year period 1992–98 in order to evaluate the adequacy of each curriculum in supporting students' learning of the basic and clinical sciences.

Methods

From 1987 to 1999, the Wake Forest University School of Medicine operated two curricular tracks; the problem-based Parallel Curriculum and the lecture-based Traditional Curriculum. The Parallel Curriculum enrolled 24 students out of each entering class of 108. Students applied for entry into the Parallel Curriculum following acceptance into the medical school. In the initial years of the program, 18 out of 108 students were enrolled in the Parallel Curriculum; this was subsequently expanded to 24. The Parallel Curriculum

emphasized small-group, student-centred, self-directed learning of basic and clinical science material. Tutorial groups of six students and two faculty facilitators provided the framework for the study of approximately 70 cases over a two-year period. Students identified learning issues that served as the stimulus for self-directed learning and small-group discussion. In the first year, basic science correlates in gross anatomy, histology, pathology, radiology and biochemistry were provided. In the first year, students met three times a week for two-hour sessions and in the second year, the PBL groups met twice for two- and half-hour sessions. Evaluations were multifaceted and included standardized patient examinations. Students in the Parallel Curriculum track had limited experience with multiple choice examinations.

The Traditional Curriculum was primarily organized on a discipline-specific basis and utilized lectures as the primary educational methodology. Major first-year courses included anatomy, biochemistry, micro-anatomy, immunology, physiology and neuroscience. Second-year courses included Pathology, Pharmacology, Medical Microbiology and 17 organ-system based courses. The Traditional Curriculum primarily utilized multiple-choice examinations for evaluation of knowledge acquisition.

Students in both pre-clinical tracks completed the same third-year clerkship schedule and were intermingled during the rotations. The third year was composed of clerkships in Surgery, Medicine, Paediatrics, Obstetrics and Gynaecology, Psychiatry, Neurology, Family Medicine, Anaesthesiology and Radiology.

Six hundred and eighty-nine students who took the United States Medical Licensing Examination Step 1 and 540 students who took the United States Medical Licensing Examination Step 2 for the first time over the seven-year period (1992–98) were included in the analyses. Wake Forest University School of Medicine (WFUSM) students take Step 1 at the end of the second year and Step 2 at the end of the third year of medical school. Individual student scores, mean Wake Forest University School of Medicine scores and national mean scores for the Step 1 and Step 2 examinations were obtained from the score reports provided to WFUSM by the the National Board of Medical Examiners. Two-way Analysis of variance (ANOVA) was used to compare the mean student performance in Step 1 and Step 2 by cohort year (1992–98) and curricular track (Traditional vs. Parallel). *T*-test analyses were also performed separately for each cohort year to compare mean student performance in Step 1 and Step 2 by curricular tract. Mean total MCAT scores and biology sub-scores for students were obtained from the school's Office of Student Admissions. *T*-test analyses were employed to compare MCAT scores of students in the two curricular tracks for the entering classes of 1990–96.

Results

Table 1 displays mean total MCAT and Biology subscores for the entering classes of 1990–96. These classes took the Step 1 exam beginning in 1992 and the Step 2 exam beginning in 1993.

Parallel Curriculum students had higher mean biology and total MCAT scores in three of the seven years studied. Statistically significant ($P = \text{or} < 0{\cdot}05$) differences between the two curricular tracks were noted in only one of the years studied (1994) for the total MCAT score.

Table 2 displays mean United States Medical Licensing Examination Step 1 scores by year and curricular type for the seven-year period 1992–98.

Table 1 MCAT scores – total and biology sub-score entering classes 1990–96

Entering Class						
		Biology Sub-score			Total MCAT	
Year	Parallel	Traditional	*P*-Value	Parallel	Traditional	*P*-Value
1990	9·65	9·40	0·52	54·57	53·86	0·66
1991	10·04	9·91	0·71	58·04	57·42	0·70
1992	8·92	9·60	0·10	26·67	28·70	0·05
1993	9·05	9·31	0·50	27·50	29·05	0·14
1994	9·04	9·56	0·26	28·52	28·74	0·85
1995	10·29	10·05	0·47	30·25	29·70	0·49
1996	9·52	10·05	0·19	29·17	29·59	0·65

Table 2 USMLE Step 1 scores Wake Forest University School of Medicine and national averages 1992–98 first-time test-takers

Cohort	1992	1993	1994	1995	1996	1997	1998
WFUSM[1]	206	213	205	213	215	217	218
Traditional	207	213	211	216	215	216	219
Parallel	201	213	188	202	218	220	213
National	201	204	205	208	210	213	216

[1]Total for WFUSM – includes all students (Traditional and Parallel).

Table 4 USMLE Step 2 scores Wake Forest University School of Medicine and national averages 1992–98 first-time test-takers

Year	1992	1993	1994	1995	1996	1997	1998
WFUSM[1]	207	205	203	207	213	210	208
Traditional	209	205	203	208	215	211	207
Parallel	199	205	206	200	206	208	215
National	200	200	199	202	207	209	209

[1]Total for all WFUSM students combined (Traditional and Parallel).

United States Medical Licensing Examination Step 1 mean scores over the seven-year period (1992–98) were 214 for Traditional Curriculum students and 208 for Parallel Curriculum students (t-value = 1·32, $P = 0·21$). Wake Forest University School of Medicine students scored above the national mean for six of the seven years studied. Mean scores of Parallel Curriculum students were equal to or greater than their Traditional Curriculum counterparts in three of the seven years studied. The statistically significant lower mean scores for Parallel Curriculum students in 1994 and 1995 reflect the disproportionately higher number of Parallel Curriculum students failing the Step 1 exam in those years. There is no apparent trend in the scores of the Parallel Curriculum students, although the mean scores in the latter three years of the Analyses (1996–98) exceed those in the first four years (1992–95). A positive trend in Traditional Curriculum student mean scores is noted across the seven-year period paralleling the increase in national mean scores. During the period 1992–98, there were 19 Parallel Curriculum students who failed Step 1 on their first attempt compared to 20 Traditional Curriculum students.

Table 3 displays mean USMLE Step 1 scores by cohort year and curricular track. Two-way ANOVAS indicate statistically significant main effects by cohort year (F-3·6, $P < 0·05$) and curricular track ($F = 4·3$, $P < 0·05$). There is also an interaction between the effects of cohort year and curricular track ($F = 6·3$, $P < 0·01$). This is exemplified by significant differences between the two curricular tracks in both 1994 and 1995 cohort years, as indicated by separate t-test analyses. There were no statistically significant differences between the two tracks in the other cohort years.

Table 4 displays mean United States Medical Licensing Examination Step 2 scores by year and curricular type for the seven-year period 1992–98.

Mean Step II scores over the seven-year period were 208 for Traditional Curriculum students and 206 for Parallel Curriculum students (t-value = 1·08, $P = 0·30$). The mean scores of Parallel Curriculum students exceeded or were equal to their Traditional Curriculum counterparts in three of the seven years studied (1993, 1994, 1998). There was variation by year in scores similar to that seen above for Step 1 scores. There is no obvious trend in the scores of Traditional Curriculum students across the seven years. Excluding 1995 scores of Parallel Curriculum students, the scores have shown a positive trend since 1992. The increase in national mean scores over this time period has not been as

Table 3 Mean differences in Step 1 student performance by cohort year and curricular track

Cohort year	Traditional mean (SD)	Parallel mean (SD)	Mean difference (P-value) NS = Not significant at $P < 0·05$
1992	207 (18)	201 (15)	6 (NS)
1993	213 (24)	213 (17)	(NS)
1994	211 (18)	188 (25)	23 ($P < 0·001$)
1995	216 (20)	202 (24)	14 ($P < 0·01$)
1996	215 (22)	218 (20)	3 (NS)
1997	216 (16)	220 (20)	4 (NS)
1998	219 (20)	213 (15)	6 (NS)

Table 5 Mean differences in Step 2 student performance by cohort year and curricular track

Cohort year	Traditional mean (SD)	Parallel mean (SD)	Mean difference (P-value) NS = Not significant at $P < 0·05$
1992	209 (21)	199 (25)	10 ($P < 0·05$)
1993	205 (17)	205 (18)	NS
1994	203 (15)	206 (18)	3 (NS)
1995	208 (22)	200 (25)	8 (NS)
1996	215 (24)	206 (17)	9 ($P < 0·05$)
1997	211 (22)	208 (20)	3 (NS)
1998	207 (20)	215 (24)	8 (NS)

pronounced as for Step 1. Eight Parallel Curriculum students failed the Step 2 exam on their first attempt compared with 13 Traditional Curriculum students during the seven-year period.

Table 5 displays mean USMLE Step 2 scores by cohort year and curricular track. Two-way ANOVAS again indicate a statistically significant main effect by curricular track ($F = 4{\cdot}1$, $P < 0{\cdot}05$). There is also an interaction between the effects of cohort year and curricular track ($F = 6{\cdot}1$, $P < 0{\cdot}01$). *T*-test analyses reveal that there are statistically significant differences in Step 2 mean student performance in cohort year 1992 and 1996. There were no statistically significant differences by curricular track in the other cohort years.

Discussion

These results support the hypothesis that the standardized test performance of students enrolled in a problem-based curriculum are comparable to that of students enrolled in a more traditional curricular track. However, students in the Traditional Curriculum had higher mean scores on both Step 1 and Step 2 across the seven-year period studied, though these differences were not statistically significant. The results are similar to previous findings by Goodman and colleagues for the National Board of Medical Examiners Part I and II at Rush Medical College but differ from those of Mennin and colleagues at the University of New Mexico for Step 1.[2,3] We did not see the increase in Step 1 scores observed by Blake and colleagues following implementation of a PBL curriculum at the University of Missouri-Columbia.[5]

The variation in mean scores between the years is also of interest. Since the Parallel Curriculum was implemented in 1987, there has been a movement toward inclusion of more directive task questions within the cases. Modification of the Traditional Curriculum also occurred over this period with an increase in small-group learning experiences. The net effect of these changes was to move the two curricular tracks away from the extremes of the curricular spectrum they once occupied. There was little change in the third-year clerkships over the same time period.

The curricular changes noted above do not, by themselves, fully explain the variation in scores between the years seen for both Step 1 and Step 2. This difference may also reflect the composition of the students within each curricular track in a given year. There are statistically significant main effects by cohort year and curricular track. Students entering with a solid basic science foundation as reflected by MCAT score tended to do better on both exams regardless of curriculum type. These results emphasize the importance of a solid basic science foundation prior to entering medical school. There were undoubtedly other student characteristics which may have played a role in influencing the ability of a student to acquire basic science knowledge within their chosen curricular track.

During this two-year interval (1992–93), there was a difference in the admissions practice for the Parallel Curriculum as compared to other years. This resulted in a higher proportion of students with poorer basic science preparation entering the Parallel Curriculum and contributed to a higher failure rate on Step 1 in 1994 and 1995 which was reflected in lower mean scores for the Parallel Curriculum cohort in these two years.

Longitudinal study of curricular outcomes provides an enhanced opportunity to evaluate programme effectiveness as compared with cross-sectional one-year evaluations. Colliver has noted that determination of the effectiveness of problem-based learning will require outcome studies examining the impact on clinical performance, preferably after graduation.[6] Santos-Gomez and colleagues analyzed the performance of PCC graduates from the University of New Mexico for the Classes of 1985–87 along eight dimensions including knowledge and independent learning ability.[7] Residency supervisors' average ratings tended to be higher for PCC graduates, but the difference was not statistically significant. Although this study examines long-term trends in standardized test scores of problem-based learning students, it does not directly link these intermediate outcomes with the clinical performance of these students. A previous study conducted by Richards *et al.* at Wake Forest University School of Medicine demonstrated no significant difference in the National Board of Medical Examiners Medicine Shelf Exam scores between Parallel Curriculum and Traditional Curriculum students, but Parallel Curriculum students were rated higher than their Traditional Curriculum counterparts on four clinical rating scales including amount of factual knowledge.[8] Distlehorst and colleagues also demonstrated higher third-year clerkship ratings for PBL cohort students over the three year period analyzed.[4]

There were no statistically significant differences between students in the two curricular tracks taken as a whole cohort across the seven-year period. In fact, as Berkson has noted, the similarity between the two groups may be real.[8] The totality of experience in both groups, although differing by curricular type, may be similar enough that the comparable scores are what should be expected. We would agree with Berkson that the focus then shifts to identifying what common elements of the curricular experience can be blended to produce the most desirable outcome.

Conclusion

This study provides evidence of the general, long-term effectiveness of the Parallel Curriculum at Wake Forest University School of Medicine with respect to basic and clinical science knowledge acquisition. These results should be encouraging to curricular planners and faculty with respect to inclusion of problem-based learning in the medical school curriculum. At Wake Forest University School of Medicine, problem-based learning was incorporated into the new *Prescription For Excellence* curriculum, which began in August 1998 for all matriculating students. The presence of the Parallel Curriculum had demonstrated the value of problem-based learning in terms of basic science knowledge acquisition and enhanced student satisfaction with the learning process.

Contributors

CE and LC-L contributed to all aspects of this paper, including initial conceptualization of the paper, data analysis, writing of the paper and revision in response to reviewer comments.

Acknowledgements

The assistance of Ms. Kimberly Speas in preparation of this manuscript is gratefully acknowledged.

Funding

There was no funding for the preparation of this manuscript external to the Wake Forest University School of Medicine or Yale University School of Medicine.

References

1 Farquhar L, Haf J, Kotabe K. Effect of two preclinical curricula on the National Board of Medical Examiners Part I examination performance. *J Med Educ* 1986;**61**:368–73.
2 Goodman LJ, Brueschke EE, Bone RC, Rose WH, Williams EJ, Paul HA. An experiment in medical education. A critical analyses using traditional criteria. *JAMA* 1991;**265**:2373–776.
3 Mennin SP, Friedman M, Skipper B, Kalishman S, Snyder J. Performance on the National Board of Medical Examiners I, II & III by medical students in the problem-based and conventional tracks at the University of New Mexico. *Academic Med* 1993;**68**:616–24.
4 Distlehorst LH, Robbs RS. A comparison of problem-based learning and standard curriculum students: Three years of retrospective data. *Teaching Learning Medicine* 1998;**10**:131–7.
5 Blake RL, Hosokawa MC, Riley SL. Student performance on Step 1 and Step 2 of the United States medical licensing examination following implementation of a problem-based learning curriculum. *Academic Med* 2000;**75**:66–70.
6 Colliver J. Research strategy for problem-based learning: cognitive science or outcomes research? *Teaching Learning Medicine* 1999;**11**:64–5.
7 Santos-Gomez L, Kalishman S, Rezler A, Skipper B, Mennin SP. Residency performance of graduates from a problem-based and a conventional curriculum. *Med Educ* 1990;**24**: 366–75.
8 Richards BF, Ober P, Cariaga-Lo, Camp MG, Philp J, McFarlane M, Rupp R, Zaccaro DJ. Ratings of students' performances in a third-year internal medicine clerkship: a comparison between problem-based and lecture-based curricula. *Academic Med* 1996;**71**:187–9.
9 Berkson L. Problem-based learning. have the expectations been met? *Academic Med* 1993;**68**:S79–88.

Received 11 *January* 2000; *editorial comments to authors* 13 *March* 2000; *accepted for publication* 1 *March* 2001