

USMLE Step 2 Performance and Test Administration Date in the Fourth Year of Medical School

CHARLES A. POHL, MARY R. ROBESON, and JON VELOSKI

ABSTRACT

Purpose. To determine whether the time interval between completing the third-year curriculum and test administration affects a student's USMLE Step 2 score.

Method. Scores for 846 students in the classes of 2000–2004 were grouped in ten time periods depending on test date. A linear regression model to predict performance on Step 2 using gender, Step 1, and grades in medicine, pediatrics and obstetrics–gynecology was developed based on the class of 1999. Analysis of covariance was used to test the effect of time on scores, adjusting for predicted performance.

Results. Step 2 scores decreased significantly ($p < .001$) across time. Students' mean scores were four points higher than predicted in the early months and five to eight points lower near the end of the senior year.

Conclusions. Students who scheduled Step 2 early in the senior year achieved higher scores, on average, than those who waited until later in the year.

uled the test in June. Similarly, April and May were combined because only ten students had delayed their test until May. A group of 179 members of the class of 2000 could not be included because they had completed Step 2 before the computerized test with flexible scheduling was available.

We recognized that there might be inherent differences in the academic ability of students that scheduled Step 2 at different times. In order to adjust for these differences, a full linear regression model was developed using data for 217 students who had graduated in 1999 before computer administration. Step 2 score on the first attempt was the dependent variable. The independent, predictor variables included Step 1 scores, gender, and grades based on departmental examinations in family medicine, medicine, pediatrics, psychiatry, surgery and obstetrics–gynecology. Gender was included because local as well as national studies have shown that women's performances in clinical disciplines exceed their performances in the preclinical disciplines, especially on the board examinations.^{3–5} The model accounted for 74% of the variance in Step 2 scores with a standard error of estimate of ± 10 . The beta weight for Step 1 was .66. The beta weights for gender, medicine, pediatrics, and obstetrics–gynecology fell in the range of .10–.15. All were significant by *t*-test at $p < .01$. The course grades in family medicine, psychiatry, and surgery were dropped from the final model because their betas were not statistically significant. Cross-validation on the class of 2000 yielded a product-moment correlation of .84 between predicted and actual Step 2 scores.

The linear regression model was used to predict Step 2 scores for the students in the graduating classes of 2000–2004 using data available before they completed the computer version of Step 2. Differences between the mean predicted and actual scores for each time period were computed. The students' actual Step 2 scores across the ten time periods were subjected to analysis of covariance with adjustment for their predicted performance on Step 2 within each time period in order to test the hypothesis of no difference among months of test administration.

Results

Approximately two-thirds of the students scheduled Step 2 sometime in June through November of their fourth year, as shown in Table 1. The largest fraction chose November. Less than one-tenth postponed their test date until March, April, or May of their senior year. The overall mean of 213, standard deviation of 22 and range (142–273) in the sample of four classes were close to national norms.

There were statistically significant ($p < .02$) differences in the students' mean predicted scores across time. The overall differences in mean actual scores adjusted for predicted scores across time periods were also statistically significant ($p < .001$). The students who completed the test in June through August achieved the largest increases as indicated by the difference between their predicted and actual scores. However, there were decreases, on average, for the students who postponed the test until later in the senior year.

The overall passing rate was 95.2%, which ranged from a high of 97.3% for those who tested in June/July to a low of 84.4% for those in March. The overall differences in pass rate across the time periods were not statistically significant (chi-square₍₉₎ = 11.7, $p < .23$).

Young physicians must pass the United States Medical Licensing Examination (USMLE™) for licensure and specialty board certification. During the 2000–01 academic year, 101 U.S. medical schools required that students attempt Step 2 at some point during their MD program and 72 schools required that the student pass the examination for their MD degree.¹ The consequences of performance on the USMLE extend beyond medical school, as Step 1 and occasionally Step 2 have become important factors in screening residency candidates through the residency application process.² Step 2 has even greater consequences among those who are at risk of failing this cumulative, high-stakes examination. That is, residency programs are less likely to consider and rank applicants who failed or are at risk of failing Step 2 because they may not graduate from medical school in time to start their graduate medical education.

Before it became available on computer in 2000, Step 2 had been administered as a paper-and-pencil test in the fall and spring of the students' fourth year of medical school. Now students can schedule the computer-based test at their own convenience throughout the year, subject to their medical school's curriculum and the faculty's policies for promotion and graduation. While little is known about the effect of this change to flexible scheduling on student performances, such information could be useful to students and faculty advisors. We designed this study to analyze the relationship between students' Step 2 scores and the time interval that had passed since they had completed the third-year medical school curriculum.

Method

Subjects included 846 students in the graduating classes of 2000–2004 who were required to pass USMLE Steps 1 and 2 at a large private medical school. The test dates for their first attempt at Step 2 were collapsed into a series of ten-month-long time periods between the end of their third year and the end of their fourth year. June and July were combined because only 32 students had sched-

TABLE 1. Performance on USMLE Step 2 by Month of Test Administration in the Fourth-Year Curriculum*

Month of Test Administration	n	Total Score on USMLE Step 2				Difference Mean
		Predicted		Actual		
		Mean	SD	Mean	SD	
June/July	105	211	15	215	29	+4
August	114	214	18	216	22	+2
September	89	211	18	212	18	+1
October	84	212	17	211	20	-1
November	146	216	20	218	22	-2
December	128	215	17	212	23	-3
January	58	213	18	209	24	-4
February	49	217	22	212	22	-5
March	39	208	18	203	22	-5
April/May	34	214	20	206	23	-8
Total	846	213	18	213	22	0

*Predicted scores based on a linear regression model ($R^2 = .74$) using Step 1 score, gender, and test grades in medicine, pediatrics, and obstetrics-gynecology. Differences in mean actual scores ($F_{9,845} = 4.96$) are statistically significant ($p < .001$) by analysis of covariance with adjustment for predicted score.

Discussion

Medical students face a dilemma as they choose a test date for Step 2. Some students may be inclined to sit for the examination as quickly as possible after completing core clerkships in order to pass the test and move unencumbered into their final year of medical school. Others may delay the examination in order to concentrate on the residency application process or to review and integrate what they have learned to assure thorough preparation for the test.⁶ The latter strategy has become even more critical for students who followed conventional, departmental-based curricula as the interdisciplinary content specifications of the new USMLE have replaced the traditional discipline-based topics of the earlier National Board Examinations.⁷ In addition, some at-risk students delay the examination date so that potential residency programs will not have access to their examination scores.

Although research in the cognitive sciences supports the significance of short time intervals in predicting retention and recall after learning,⁸ the literature on test-taking provides conflicting evidence that medical school faculty can use to advise students on their choice of test date. One study of the timing of students' Step 1 test date concluded "the later, the better." However, this study was limited by its small number of subjects, use of sample questions rather than a live USMLE test, and a high failure rate.⁹ A more recent study, which involved over 600 subjects who scheduled Step 1 over a two-month time span at the end of the second year of medical school, concluded that the timing of a comprehensive medical examination did not affect performance.¹⁰ In addition, earlier studies of the timing between completing certain core clerkships and later performance on Part II of the National Board Examination concluded that shorter intervals between learning and testing led to higher scores in obstetrics-gynecology, psychiatry, and surgery.¹¹⁻¹⁴

The findings of the present study have significant implications for those who advise students about scheduling the examination. Students, on average, benefited most by taking the test during the early months of their senior year. Therefore, students should not be advised to delay taking the test under the assumption that this strategy will enhance their scores. The results indicate that students are not at a disadvantage if, after proper test preparation, they sit for the exam in early June or July immediately after completing their third-year curriculum. Although the findings of the present study are consistent with earlier studies of Part II that found higher test scores with shorter time intervals, faculty advisors must also con-

sider that students' career interests influence performance on clinical examinations such as Part II or Step 2.¹⁵⁻¹⁷

A possible criticism of this study is that a student's ability to pass the comprehensive test is more important than the numerical score because the residency decision may not hinge on the Step 2 score. However, given that the majority of medical schools require a passing score on Step 2 for graduation, the examination is a high-stakes test, especially for students at risk of failing. As such, most residency programs require a passing Step 2 score among those students who have previously failed or nearly failed USMLE Step 1 in order to be considered as a viable residency candidate.

Students' test scores may have been influenced by other factors beyond the timing of the examination. Test preparation including study time, educational materials, and coaching courses may have influenced performance. Previous studies of these factors, especially coaching courses, reported no effect on USMLE Step 1 scores.⁶ A student's study behavior (e.g., procrastination, enthusiastic preparation) also may have affected the test score. In addition, some students may have been distracted by personal or health issues during the study period. Future investigation is necessary to account for these individual variables in order to counsel students about specific methods of preparing for the comprehensive examination.

This study used a quasi-experimental design in which analysis of covariance revealed systematic differences in Step 2 scores across testing dates adjusting for differences in gender and prior academic performance. The linear regression model using a set of five independent variables explained 74% of the variance in the dependent variable, Step 2. The differences in predicted scores across time periods implied some self-selection confounded with academic ability. While statistical adjustment without true randomization has limitations, it was impractical to consider random assignment of students' testing schedules in order to avoid this design. A larger sample with a higher failure rate distributed more evenly across test dates is needed to address this question.

We conclude that the time interval between the end of the third-year medical curriculum and the date of the USMLE Step 2 examination does affect students' scores. The mean differences are consistent with other studies and may be critical if the test-taker is at risk of poor performance. These results have important implications for counseling students about their scheduling decisions and underscore an opportunity for further study of the relationship between students' background, test preparation, and scheduling for high-stakes examinations.

Correspondence: Charles A. Pohl, MD, Jefferson Medical College of Thomas Jefferson University, 157 Alumni Hall, 1020 Locust Street, Philadelphia, PA 19107-6799; e-mail: (charles.pohl@jefferson.edu).

References

1. Barzansky B, Etzel SI. Educational programs in US medical schools, 2000-2001. *JAMA*. 2001;286:1049-55.
2. Adams LJ, Brandenburg S, Lin C, Blake M, Lemenager M. National survey of internal medicine residency programs of their 1st-year experience with the electronic residency application service and national resident match program changes. *Teach Learn Med*. 2001;13:221-6.
3. Dawson B, Iwamoto CK, Ross LP, Nungester RJ, Swanson DB, Volle RL. Performance on the National Board of Medical Examiners Part I examination by men and women of different race and ethnicity. *JAMA*. 1994;272:674-9.
4. Arnold L, Willoughby TL, Calkins V, Jensen T. The achievement of men and women in medical school. *J Am Med Womens Assoc*. 1981;36:213-21.
5. Herman MW, Veloski JJ. Premedical training, personal characteristics and performance in medical school. *Med Educ*. 1981;15:363-7.
6. Thadani RA, Swanson DB, Galbraith RM. A preliminary analysis of different approaches to preparing for the USMLE Step 1. *Acad Med*. 2000;75:S40-2.
7. Swanson DB, Case SM. Assessment in basic science instruction: Directions for practice and research. *Adv Health Sci Educ*. 1997;2:71-84.
8. Regehr G, Norman GR. Issues in cognitive psychology: implications for professional education. *Acad Med*. 1996;71:988-1001.
9. Petrusa ER, Reilly CG, Lee LS. Later is better: projected USMLE performance during medical school. *Teach Learn Med*. 1995;7:163-7.

10. Pohl CA, Robeson M, Hojat M, Veloski JJ. Sooner or later? USMLE Step 1 performance and test administration date at the end of the second year. *Acad Med.* 2002;77:S17-9.
11. Hojat M, Veloski JJ. Subtest scores of a comprehensive examination of medical knowledge as a function of retention interval. *Psychol Rep.* 1984;55:579-85.
12. Ripkey DR, Case SM, Swanson DB. Predicting performances on the NBME surgery subject test and USMLE Step 2: the effects of surgery clerkship timing and length. *Acad Med.* 1997;72:S31-3.
13. Ripkey DR, Case SM, Swanson DB. Identifying students at risk for poor performance on the USMLE Step 2. *Acad Med.* 1999;74:S45-8.
14. Smith ER, Anderson G. A decrease from 8 to 6 weeks in obstetrics and gynecology clerkships: effect on medical students' knowledge. *Obstet Gynecol.* 2002;86:458-60.
15. Gonnella JS, Hojat M, Erdmann JB, Veloski JJ. The impact of early career specialization on licensing requirements and related educational implications. *Adv Health Sci Educ.* 1997;1:125-39.
16. Gonnella JS, Veloski JJ. The impact of early specialization on the clinical competence of residents. *N Engl J Med.* 1982;306:275-7.
17. Williams T, Sachs L, Veloski JJ. Performance on the NBME Part II examination and career choice. *J Med Educ.* 1986;61:979-81.