

Study of Selected Outcomes of Medical Students Who Fail USMLE Step 1

Diane M. Biskobing, MD*, Sonya R. Lawson, PhD†, James M. Messmer, MD, MEd‡,
J. Dennis Hoban, EdD†

*Department of Medicine and Dean's Office

†Dean's Office

‡Department of Radiology and Dean's Office

Virginia Commonwealth University

School of Medicine

Richmond, VA

Abstract

Purpose: Passage of the United States Medical Licensing Exam (USMLE) is required to obtain a medical license in the United States. Currently the majority of US medical schools require passage of USMLE Step 1 for either promotion to the third year or graduation from medical school. Virginia Commonwealth University School of Medicine (VCUSOM) requires that students take the USMLE but does not require passing of USMLE Step 1 for promotion or graduation. This policy enabled the authors to analyze performance outcomes during clinical rotations and monitor the residency match for a group of students who failed USMLE Step 1 on the first attempt.

Methods: Third year clerkship grades and residency match results were reviewed for 64 students of the graduating classes of 1999-2005 who failed Step 1 on the first attempt. An equal number of students who passed Step 1 were randomly selected from each class as a comparison group. Average clinical performance ratings, NBME subject exam scores and final third year clerkship grades for the two groups were compared. Residency match rates and specialty certification were also compared.

Results: The USMLE Fail Group had more Pass and fewer Honors clerkship grades than the comparison group. Subject exam scores were significantly lower in the USMLE Fail Group in all clerkships. Clinical performance ratings were significantly lower in the Fail group in three out of six clerkships: Internal Medicine, Pediatrics, and Psychiatry. However, 82% of the USMLE Fail Group later passed USMLE Step 1 and 2. Fifty-nine of the 64 students in the USMLE Fail Group matched for a residency, whereas all of the students in the Pass Group matched for a residency.

Conclusion: Students who fail USMLE Step 1 have lower final clerkship grades due in part to lower NBME subject exam scores. The majority of these students, however, successfully pass USMLE Step 1 prior to graduation, go on to graduate medical training, and become board certified in their specialty.

Passage of the United States Medical Licensing Exam (USMLE) is required to obtain a medical license in the United States. USMLE Step 1 is ordinarily taken between the second and third year of medical school. In 2005, 114 US medical schools required passage of USMLE Step 1 prior to graduation and 96 of those schools required passage for promotion to the third year clerkships.¹ VCUSOM was one of only 11 US medical schools that required students take the USMLE exam, but did not require passage for promotion or graduation. VCUSOM students who fail USMLE Step 1 on the first attempt either take time off during the third year clerkships to retake the exam or defer it to the fourth year. This policy of not requiring passage of USMLE Step 1 prior to graduation allowed us to review the following

outcomes in students who fail USMLE Step 1 on the first attempt: performance in the third year clerkships, residency match results, and board certification.

The ability of a basic science examination to predict performance on clinical examinations has been studied extensively. Prior to the development of the USMLE, performance on the National Board of Medical Examiners (NBME) Part I examination was shown to moderately correlate with NBME Part III scores ($r=0.56$).² More significantly, performance on NBME Part I somewhat predicted performance on certifying examinations in Internal Medicine, Orthopedic Surgery Preventive Medicine, and Dermatology ($r = .42-.63$).³

Whether the USMLE Step 1 examination predicts actual clinical performance is a separate issue. Investigators have reported mixed results. Fields and others⁴ reported that performance in the first two years of medical school and USMLE Step 1 scores were related to third year clerkship GPA. Performance in the Principles of Clinical Medicine course, however, showed the strongest correlation with third year GPA. Roop and Pangaro⁵ similarly found that preclerkship GPA accounted for 28% of the variance of the medicine clerkship grade. They did not evaluate the relationship with USMLE scores. The relationship of NBME Part 1 scores with residency performance has also been studied. NBME Part I was reported to have a weak relationship with clinical ratings during Internal Medicine residency.⁶ Other studies reported no relationship between NBME Part I scores and faculty clinical ratings of residents.^{7,8} More recently the relationship between USMLE scores and performance in residency has been evaluated. Paolo and others⁹ reported a correlation between both USMLE Step 1 and 2 scores and residency director's ratings. Andriole and others,¹⁰ as well, reported a modest correlation between USMLE Step 2 scores and surgical interns ratings by program directors but found no relationship with USMLE Step 1 scores.

Given the unique situation at VCUSOM in which students are required to take but not pass USMLE Step 1 prior to starting their clinical clerkships, we undertook the following study. We compared the performance of third year clerks who failed USMLE Step 1 on the first attempt and a group who passed it on the first try. Since clerkship grades are a composite of both multiple-choice examination scores and faculty and residents' clinical ratings we were able to examine each component of the final grade for these two groups. None of the previous studies investigated the specific components of clerkship grades. We also compared residency match rates and board certification in the subset of the groups who had completed Internal Medicine, Pediatrics, or Family Medicine training.

Methods

Sixty-six VCUSOM students of the graduating classes of 1999 through 2005 failed USMLE Step 1 on the first attempt. We analyzed third year clerkship grades and match results for 64 of these students (Fail Group) and a randomly chosen comparison group of an equal number of students picked from each class who passed Step 1 (Pass Group). The Pass Group was obtained using a random numbers table. Since we had to review individual student files to obtain all the information presented, we did not compare the Fail Group to the entire class. This number of Fail students constituted 5.7% of the 1163 to-

tal students from the classes of 1999 through 2005. We reviewed clerkship grades only on the 64 students who graduated from medical school. Two students who had failed Step 1 were dismissed from school during the third year for academic reasons and were not included in the analysis of clerkship grades since complete data was not available. In addition to examining clerkship grades, we analyzed NBME subject exam scores on the following clerkships: Internal Medicine, Surgery, Pediatrics, Neurology, and Obstetrics-Gynecology. The Psychiatry and Family Medicine clerkships do not use subject exams. We compared clinical performance ratings for the Internal Medicine, Neurology, Obstetrics-Gynecology, Pediatrics, Family Practice, and Psychiatry clerkships. The Surgery clerkship does not use numerical scores for the clinical grade so could not be included in the analysis. The grades and NBME Subject exam scores were entered into a database without any identifying factors. Grades at our institution are based on a five point grading system of Honors, High Pass, Pass, Marginal, and Fail. Finally, USMLE Step 2 scores and residency match results were also entered into the database for student group. For the students in the classes of 1999, 2000, and 2001 that had completed Internal Medicine, Family Medicine or Pediatrics residency we also evaluated licensure and board certification status within two years of completion of residency. Statistical analysis using a t-test was performed on the subject exam and clinical ratings to determine whether differences between the group means were statistically significant. Effect size was calculated using Cohen's *d* to estimate the strength of our results (small effect = 0.2-0.4, moderate effect = 0.5-0.7, and large effect > 0.8). The study was approved by the VCU Institutional Review Board. No funding was provided for this study.

Results

By design, the Fail Group had significantly lower Step 1 scores than the Pass Group: 170 ± 9 (SD) versus 220 ± 17 ($t=20.48$, $p<0.0001$). In the Fail Group, 37 students scored between 170 and 181, 20 students between 160 and 169, and six students less than 160. USMLE Step 2 scores in the Fail Group were also significantly lower than the Pass Group: 190 ± 17 versus 226 ± 20 ($t=7.95$, $p<0.001$). In the Fail Group, 19 students failed Step 2 on their first attempt. Two of the USMLE Pass Group failed Step 2 on the first attempt but passed prior to graduation. Of the 64 students who failed Step 1 on the first try, 17 students had not passed one or both parts of USMLE by graduation ($p<0.001$): six had not yet passed Step 1, four had not passed Step 2, and seven had not passed either. Seven students in the Fail Group passed USMLE Step 1 and/or 2 after graduation indicating that 82% of the USMLE Fail group eventually passed

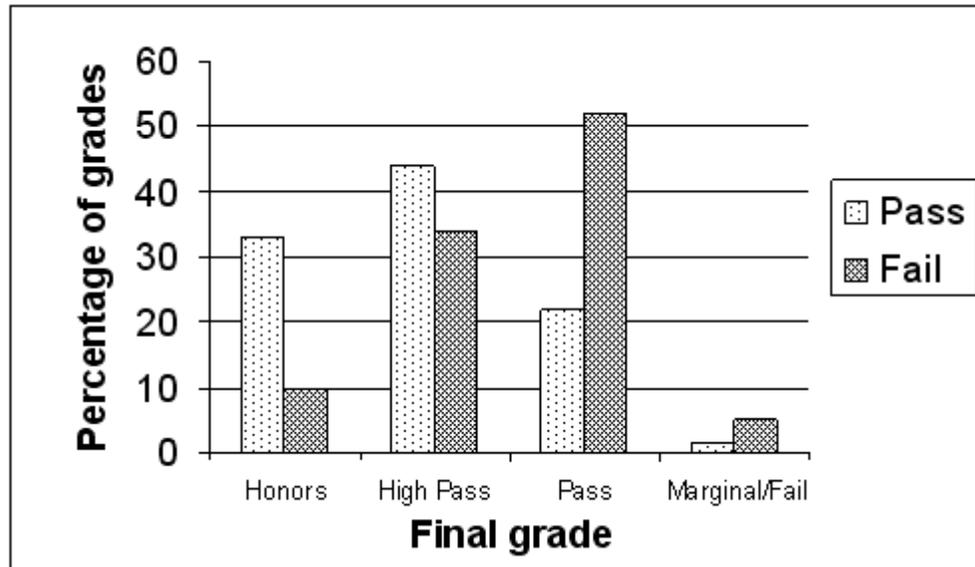


Figure 1: Distribution of final clerkship grades for USMLE Pass Group (n= 64) compared to USMLE Fail Group (n=64). The figure shows the percentage of students in each group that received Honors, High Pass, Pass, and Marginal grades for the third year clerkships. The USMLE Pass Group had more Honors and High Pass grades; USMLE Fail Group had more Pass grades and few Honors.

USMLE Step 1 and 2. However, these data may be incomplete since students can elect not to release USMLE scores to the school after graduation.

The USMLE Fail Group had more grades of Pass and Marginal and fewer grades of Honors than the USMLE Pass Group. Figure 1 shows the percentage of clerkship grades by grade category and group. A Pass grade was the most common mark (52%) in the USMLE Fail Group compared to High Pass (44%) being most common in the USMLE Pass Group. In the Fail Group, 17 students received a total of 22 Marginal grades, whereas, in the Pass Group, five students received five Marginal and one Fail grade.

Since final clerkship grades are determined from both written NBME subject exam scores as well as clinical performance ratings, we looked at each component of the grade to try to determine why the USMLE Fail Group had fewer High Pass and Honors grades. The written exam counts for a significant percentage of the final grade in each clerkship. However, each clerkship weights the exams differently. In addition, each clerkship requires a minimal passing score on the subject exam to pass the clerkship. Table 1 shows the average scores on the NBME subject exams for each group and the weighting of the exams and minimum passing score for the clerkships that use NBME subject exams. All NBME subject exams were designed to have a mean of 70 and standard deviation of eight. The USMLE Fail Group had

significantly lower exam scores than the Pass Group in all subjects. Cohen's d revealed a high level of practical significance ($d= 1.25-1.42$). We then evaluated the clinical performance ratings given by faculty and residents. This rating is a more subjective grade based on observation of performance on the wards or in outpatient clinics. The final clinical performance rating on most clerkships is a compilation of evaluations by multiple faculty attendings and residents. Table 2 lists the average clinical performance ratings for each group in six different clerkships. Each clerkship uses its own method of rating so ratings across clerkships are not comparable. Clinical performance ratings were not significantly different between groups in three out of six clerkships: Neurology, Obstetrics-Gynecology, and Family Practice. They were, however, significantly lower in the USMLE Fail Group in the Medicine, Psychiatry and Pediatrics clerkships. Cohen's d reveals a small level of practical significance between groups in Psychiatry ($d=0.4$) but a stronger impact in Medicine ($d=0.69$) and Pediatrics ($d=0.77$).

Finally, we looked at the results of the residency match for students in these two groups. In both cases, the largest number of students matched in Internal Medicine. In the USMLE Fail Group, a total of five students had not matched for a residency at the time of graduation. All of these students had not passed one or both parts of USMLE by graduation. All of the USMLE Pass Group matched for a residency. The American Boards of Internal Medicine, Pediatrics and Family Medicine publish on their websites all physicians who are board certi-

Table 1:
Subject Exam Score by USMLE Step 1 Result

Clerkship	Subject Exam (% final grade/ minimum passing score)	USMLE result					
		Pass Group (n=64)		Fail Group (n=64)		t	Cohen's d
		M	SD	M	SD		
Medicine	30%/60	76.33	9.10	66.76	5.90	-7.04*	1.25
Neurology	40%/58	75.29	6.70	66.82	5.90	-7.20*	1.34
OB/GYN	25%/59	76.88	7.80	68.98	7.00	-5.81*	1.36
Surgery	15%/62	76.65	7.30	66.71	5.70	-7.82*	1.52
Pediatrics	35%/60	74.90	7.30	66.73	4.80	-7.43*	1.32

d = 0.2-0.49 -small effect; d = 0.5-0.79 - moderate effect; d ≥ 0.8 - large effect

* p < 0.001

Of the 15 USMLE Fail Group students who have completed Internal Medicine, Pediatrics or Family Medicine residency, 11 were board certified within two years of completing residency. Of the 11 students who have completed one of these residencies and were in the USMLE Pass Group, 10 were board certified two years

Table 2
Clinical Performance Ratings by USMLE Step 1 Result

Clinical ratings	Clinical Ratings (% final grade)	USMLE result					
		Pass Group (n=64)		Fail Group (n=64)		t	Cohen's d
		M	SD	M	SD		
Medicine (54 point scale)	55%	48.51	3	46.43	3	-3.9*	0.69
Neurology (100 point scale)	50%	90.3	5	89.8	3.7	-0.54	0.11
OB/GYN (100 point scale)	75%	86	6.7	84.1	7.5	-1.48	0.27
Psychiatry (4 point scale)	35%	3.44	0.32	3.3	0.4	-2.15†	0.39
Pediatrics (10 point scale)	65%	8.92	0.4	8.58	0.48	-4.35*	0.77
Family medicine (10 point scale)	40%	8.74	0.82	8.62	0.71	-0.75	0.16

Note: Each clerkship uses their own rating scale for clinical performance.

d = 0.2-0.49 -small effect; d = 0.5-0.79 - moderate effect; d > 0.8 - large effect

Comparison of fail versus Pass Group in each clerkship

* p<0.001

† p=0.034

Table 3
Comparison of Rates of USMLE Passage, Residency Match and Board Certification

	Pass Group	Fail Group
Pass USMLE Step 1 and 2 before graduation	100%	71%
Match for Residency	100%	89%
Pass USMLE Step 1 and 2 before or after graduation	NA	82%
Medical license	91% (n=11)	87% (n=15)
Board certified in specialty	91% (n=11)	73% (n=15)

Match and USMLE passage rate are for the entire cohort (n=64). Licensure and board certification are only for those in each group who have completed an Internal Medicine, Pediatrics or Family Medicine residency.

after completion of residency. Using the available data, Table 3 summarizes the percentage of students in each group who passed USMLE Step 1 and 2, and matched for a residency. In addition, licensure data^{14,15} and board certification rates are given for those in each group who have completed Internal Medicine, Pediatrics or Family Medicine training. Licensure and board certification data may be limited by name changes or change in specialty after the initial residency match.

Discussion

Our results suggest that VCUSOM students who fail USMLE Step 1 will score lower on subject exams during the third year than those who pass Step 1. The low exam scores tend to lower final grades in the clerkships, reflected in the higher proportion of students with Pass grades in the USMLE Fail Group because the NBME subject exam score accounts for a significant portion of clerkship grades. These findings are, however, consistent with a report by Fields and others⁴ that third year GPA was related to USMLE Step 1 score and with Case and Swanson's³ observation that students who do well on past exams will continue to do well on exams. Whether the lower exam scores are a reflection of differences in medical knowledge, test taking ability, or reading skills cannot be determined from our data. Recently, however, Haught and Walls¹⁶ suggested that reading ability might be a factor in medical school success. They reported that the Nelson-Denny Reading Test subscore for Vocabulary and the Total Score were positively correlated with the USMLE Step 1 score. In a stepwise regression analysis, they determined that the vocabulary subscore was a significant predictor.¹⁶ Perhaps this should not be surprising given the fact that Ripkey and Case reported 75% of the

students who answered their post Step 2 survey used to assess performance attribution indicated that reading was perceived to affect performance, sometimes helping other times hurting it.¹⁷ It has been our experience that reading and test taking ability do play a role in academic performance as do knowledge of content and general ability. We suspect that any or all of these factors played a role in both the Fail and Pass groups in our study.

Interestingly, the clinical rating portion of the final clerkship grade was not significantly different between groups in three of the six clerkships. This may indicate that board exam scores do not predict performance in the clinical setting but rather could be a measure of test taking or reading skills. Alternatively, it may reflect the difficulty in obtaining a valid measure of a student's performance in the clinical setting. Kreiter and others¹⁸ reported that the validity of a standard clinical evaluation form increases with the number of raters for each clerkship. At our institution third year students spend the greatest clerkship time on Internal Medicine. During the 12 weeks of the Internal Medicine clerkship they are evaluated by at least four faculty and two residents. On Pediatrics, students are also evaluated by at least four different raters in several different settings over eight weeks. Psychiatry has anywhere from two to eight raters depending on the clerkship site. On other clerkships there are generally fewer number of raters and the time on each clerkship is eight weeks or less. The fact that there is a statistically significant difference in clinical performance ratings between the two groups in the Medicine, Pediatrics, and Psychiatry clerkship may reflect a greater number of evaluators resulting in better discrimination between individual students. The effect size analysis tends to support this conclusion.

Studies have found that clinical ratings by faculty during clerkships correlate with performance on USMLE Step 2 and 3,^{19,20} suggesting that clinical ratings may be valid measures of competence. Interestingly, clinical ratings during clerkships correlated with ratings of postgraduate clinical competence by program directors.^{9,19,21} Specifically, Callahan and others¹⁹ reported clinical competence ratings in Family Medicine, Internal Medicine, and Pediatrics significantly and consistently predicted residency performance. The ratings in the Internal Medicine clerkship were the best predictor of performance. Gunzburger and others²² also reported that clerkship grades significantly predicted competence as residents. In their study, third and fourth year Medicine clerkship performance, as well as performance during the Surgery clerkship, was the strongest predictors of residency performance. Markert² and Paolo and Bonaminio⁹ also reported a moderate correlation between clerkship GPA and postgraduate ratings. In the Paolo and Bonaminio⁹ study USMLE Step 1 score was found to correlate with postgraduate ratings, as well.

The most difficult task for a residency program director is to predict which applicants will perform well as residents. There have been several studies evaluating the ability of ratings during medical school to accurately predict outcomes during residency. Prior to the development of the USMLE, NBME exams were used as a measure of medical school performance. Several studies have shown a relationship between performance on NBME Part I and specialty certifying exams.^{3,6} However, the ability to perform well on standardized tests does not always predict performance in the clinical setting. Studies analyzing the ability of NBME or USMLE basic science exams to predict performance as a resident have been mixed. Both Markert² and Paola and Bonaminio,⁹ in the previously described studies, showed a correlation between NBME Part I or USMLE Step 1 scores and residency directors' ratings. However, other studies have shown no relationship between USMLE Step 1 or NBME Part I scores and clinical ratings of residents.^{7,8,10} The divergence between basic science exam scores and clinical ratings by residency directors is likely because residency directors' performance ratings are a reflection not just of medical knowledge but also other important aspects of patient care including communication, empathy, teamwork, professionalism, and leadership. These characteristics can not be ascertained from a multiple choice basic science exam.

A concern from our findings is the match results in the USMLE Fail Group. A small minority of the USMLE Fail Group did not pass either Step 1 and/or Step 2 by graduation. Of the 17 students who had not successfully

passed Step 1 or 2 by graduation, five did not match in a residency program. For some students, this was true one to two years after graduation and after participation in more than one match cycle. Other students never participated in the match for unknown reasons. We have identified a high risk group of students who may be unable to pass USMLE and become licensed physicians. It is beyond the scope of this study to determine long term outcomes in this group.

In summary, because our institution does not require passage of USMLE, we were able to analyze third year clerkship performance and residency match results in a group of students who failed USMLE Step 1 on the first attempt. Our findings indicate that initial failure of USMLE Step 1 does not foretell overall poor clinical performance during third year clerkships at our institution. However, many students who fail USMLE Step 1 on the first attempt do continue to have difficulties on standardized exams. On average, they score lower on subject exams and more often fail Step 2. As a result of lower exam scores, these students generally have lower final clerkship grades. We found a small minority of students did not pass USMLE Step 1 or 2 by graduation and did not go on to enter graduate medical education. Yet, most students (82%) successfully complete medical school, eventually pass USMLE Step 1 and 2, and match for a residency. In addition, we have shown evidence that the majority of the USMLE Fail Group go on to become board certified in their chosen specialty. This is the only published data on outcomes in students who fail USMLE Step 1 on the first attempt. USMLE Step 1 performance is just one of many variables that may predict success during third year clerkships and postgraduate training. However, whether it is predictive of performance in the long term practice of medicine is not known. The new ACGME shift toward measurement of patient outcomes during residency may help answer this question in the future.

References

1. AAMC Medical School Curriculum Directory: Institutional Characteristics [database on the Internet]. Washington, D.C. : Association of American Medical Colleges. c1995-2006 [cited 2005 Dec 7]. Available from: <http://services.aamc.org/currdir/section1/start.cfm>
2. Markert RJ. The relationship of academic measures in medical school to performance after graduation. *Acad Med.* 1993; 68 (2 Suppl):S31-4.
3. Case SM, Swanson DB. Validity of NBME Part I and Part II scores for selection of residents in

- orthopedic surgery, dermatology, and preventive medicine. *Acad Med.* 1993; 68 (2 Suppl): S51-6.
4. Fields SA, Morris C, Toffler WL, Keenan EJ. Early identification of students at risk for poor academic performance in clinical clerkships. *Acad Med.* 2000; 75(10 Suppl): S78-80.
 5. Roop SA and Pangaro L. Effect of clinical teaching on student performance during a medicine clerkship. *Am J Med.* 2001; 110:205-9.
 6. Sosenko J, Stekel KW, Soto R, Gelbard M. NBME Examination Part I as a predictor of clinical and ABIM certifying examination performances. *J Gen Intern Med.* 1993; 8:86-8.
 7. Borowitz SM, Saulsbury FT, Wilson WG. Information collected during the residency match process does not predict clinical performance. *Arch Pediatr Adolesc Med.* 2000; 154:256-60.
 8. Smith SR. Correlations between graduates' performances as first-year residents and their performances as medical students. *Acad Med.* 1993; 8:633-4.
 9. Paolo AM and Bonaminio GA. Measuring outcomes of undergraduate medical education: residency directors' ratings of first-year residents. *Acad Med.* 2003; 78:90-5.
 10. Andriole DA, Jeffe DB, Whelan AJ. What predicts surgical internship performance? *Am J Surg.* 2004; 188:161-4.
 11. Who is Certified Directory. [database on the Internet]. Philadelphia (PA): American Board of Internal Medicine. c2004-2006 [cited 2006 May 3]. Available from: <http://www.abim.org/who/index.shtm>
 12. Verification of Certification. [database on the Internet]. Chapel Hill (NC): c2003 [cited 2006 May 3]. Available from: <http://www.abp.org/>
 13. Directory of Diplomats. [database on the Internet]. Lexington (KY): American Board of Family Medicine. c2006 [cited 2006 May 3]. Available from: <https://www.theabfm.org/diplomate/find.aspx>
 14. Virginia Board of Medicine Practitioner Information. [database on the Internet]. Richmond (VA): Virginia Board of Medicine. c2002 [cited 2006 May 3]. Available from: <http://www.vahealthprovider.com/search.asp>
 15. DocFinder. [database on the Internet]. Administrators in Medicine. c1997-2005 [cited 2006 May 3]. Available from: <http://www.docboard.org/docfinder.html>
 16. Haight P, Walls R. Relationship of reading, MCAT, and USMLE Step 1 test results for medical students. *Read Psychol.* 2004; 25: 83-92.
 17. Ripkey DR, Case SM. Examinees' perceptions of factors influencing their performance on USMLE Step 2. *Acad Med.* 1996; 71 (1 Suppl): S34-6.
 18. Kreiter CD, Ferguson K, Lee WC, Brennan RL, Densen P. A generalizability study of a new standardized rating form used to evaluate students' clinical clerkship performances. *Acad Med.* 1998; 73:1294-8.
 19. Callahan CA, Erdmann JB, Hojat M, Veloski JJ, Rattner S, Nasca TJ, Gonnella JS. Validity of faculty ratings of students' clinical competence in core clerkships in relation scores on licensing examinations and supervisors' ratings in residency. *Acad Med.* 2000; 75 (10 Suppl):S71-3.
 20. Case SM, Ripkey DR, Swanson DB. The relationship between clinical science performance in 20 medical schools and performance on Step 2 of the USMLE licensing examination. *Acad Med* 1996, 71 (1 suppl): S31-S33.
 21. Hojat M, Gonnella JS, Veloski JJ, Erdmann JB. Is the glass half full or half empty? A reexamination of the association between assessment measures during medical school and clinical competence after graduation. *Acad Med.* 1993; 68 (2 Suppl): S69-76.
 22. Gunzburger LK, Frazier RG, Yang LM, Rainey ML, Wronski T. Premedical and medical school performance in predicting first-year residency performance. *J Med Educ.* 1987; 62:379-84.

Correspondence

Diane M. Biskobing, MD
PO Box 980111
Richmond, VA 23298-0111
dmbiskob@hsc.vcu.edu.