Associations Between United States Medical Licensing Examination (USMLE) and Internal Medicine In-Training Examination (IM-ITE) Scores

Furman S. McDonald, M.D., M.P.H.¹, Scott L. Zeger, Ph.D.², and Joseph C. Kolars, M.D.³

¹Residency Program Director and Associate Chair, Department of Medicine, Mayo Clinic College of Medicine, Rochester, MN, USA; ²Professor of Biostatistics and Epidemiology, Chair, Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA; ³Emeritus Program Director and Associate Chair, Department of Medicine, Professor of Medicine, Mayo Clinic College of Medicine, Rochester, MN, USA.

BACKGROUND: Little is known about the associations of previous standardized examination scores with scores on subsequent standardized examinations used to assess medical knowledge in internal medicine residencies.

OBJECTIVE: To examine associations of previous standardized test scores on subsequent standardized test scores.

DESIGN: Retrospective cohort study.

PARTICIPANTS: One hundred ninety-five internal medicine residents.

METHODS: Bivariate associations of United States Medical Licensing Examination (USMLE) Steps and Internal Medicine In-Training Examination (IM-ITE) scores were determined. Random effects analysis adjusting for repeated administrations of the IM-ITE and other variables known or hypothesized to affect IM-ITE score allowed for discrimination of associations of individual USMLE Step scores on IM-ITE scores.

RESULTS: In bivariate associations, USMLE scores explained 17% to 27% of the variance in IM-ITE scores, and previous IM-ITE scores explained 66% of the variance in subsequent IM-ITE scores. Regression coefficients (95% CI) for adjusted associations of each USMLE Step with IM-ITE scores were USMLE-1 0.19 (0.12, 0.27), USMLE-2 0.23 (0.17, 0.30), and USMLE-3 0.19 (0.09, 0.29).

CONCLUSIONS: No single USMLE Step is more strongly associated with IM-ITE scores than the others. Because previous IM-ITE scores are strongly associated with subsequent IM-ITE scores, appropriate modeling, such as random effects methods, should be used to account for previous IM-ITE administrations in studies for which IM-ITE score is an outcome.

KEY WORDS: United States Medical Licensing Examination (USMLE); Internal Medicine In-Training Examination (IM-ITE); medical knowledge; assessment; random effects modeling.

J Gen Intern Med 23(7):1016-9

DOI: 10.1007/s11606-008-0641-x

© Society of General Internal Medicine 2008

INTRODUCTION

Medical knowledge is one of the six competencies required of residents by the Accreditation Council for Graduate Medical Education (ACGME).¹ Standardized exams are well suited to the assessment of medical knowledge. The Internal Medicine In-Training Examination (IM-ITE) is given each October to over 18,000 residents enrolled in internal medicine residencies in the US.² This validated multiple choice examination was developed jointly by the American College of Physicians (ACP), the Association of Program Directors of Internal Medicine (APDIM), and the Association of Professors of Medicine (APM).³ Minimum scores on the IM-ITE are associated with passing scores on the American Board of Internal Medicine Certifying Examination (ABIM-CE).⁴⁻⁸

The United States Medical Licensing Examination (USMLE) was developed by the Federation of State Medical Boards (FSMB) and the National Board of Medical Examiners (NBME).⁹ These well-validated multiple choice examinations are given in three steps designed to assess skills for basic sciences, supervised medical care, and medical care without supervision, respectively.⁹ Although the exact timing can vary, trainees usually take USMLE Step 1 (USMLE-1) after the 2nd year of medical school with Step 2 (USMLE-2) taken in the 4th year of medical school,¹⁰ and Step 3 (USMLE-3) taken during the 1st year of post-graduate training. Passing all three steps of the USMLE is usually required for medical licensure in the US.^{9,11,12}

Many medical education research studies have used scores on these standardized examinations as outcome measures.¹³⁻²⁰ The associations of previous USMLE Step scores on subsequent USMLE Step scores and IM-ITE scores are important to correctly account for these examinations in studies of educational outcomes. The purpose of this analysis was to evaluate these associations.

METHODS

Subjects and Study Design

The Internal Medicine Residency, Mayo Clinic, Rochester, MN, trains approximately 48 residents per class in three sequential years of categorical internal medicine residency. All 195 categorical internal medicine residents enrolled in this resi-

Table 1. Number of Residents Per Class Who Took the Internal Medicine In-Training Examination (IM-ITE) During Each Year*

Year of IM-ITE administration	Class of 2003 (n=48)	Class of 2004 (n=49)	Class of 2005 (n=52)	Class of 2006 (n=46)	Total IM-ITE administrations
2000	44	–	–	–	44
2001	48	44	–	–	92
2002	47	48	48	–	143
2003	–	49	47	46	142
All years					421

*All 195 residents were encouraged to take the exam each year, but not all could do so due to absences (e.g., away elective, illness, transfer)

dency during the 2002 and 2003 administration of the IM-ITE formed the cohort for this study. This cohort was identified for a previous study designed to assess associations of medical knowledge acquisition with IM-ITE scores.¹⁸ Because the IM-ITE is given yearly, a total of 421 IM-ITE scores from the years 2000 through 2003 were available for this analysis (Table 1). USMLE-1 scores were available for all 195 residents who took the IM-ITE. Due to the timing of the other USMLE Steps, scores for these examinations were not available for all residents. USMLE-2 scores were available for 186 residents, and USMLE-3 scores were available for 89 residents (Table 2).

Data Analysis

Graphic and tabular displays were used to assess the data for patterns and trends. Linear regression techniques were used to assess the associations between USMLE Step scores. Because the multiple scores available for each resident on the IM-ITE were likely autocorrelated, a random effects model was used for assessment of IM-ITE score as a dependent variable.^{18,21,22} This model also allowed for adjustment for multiple known and hypothesized predictors of IM-ITE score including: conference attendance in the year prior to the IM-ITE, use of *UpToDate* in the year prior to the IM-ITE, post-graduate year (PGY), United States vs. International Medical Graduate status (USMG vs IMG), National Residency Matching Program rank (NRMP), Objective Structured Clinical Examination score (test of physical examination skill), class rank (ordinal ranking based on residency rotation grades), age at entry into residency, gender, marital status, and parenthood (whether or not a resident had a child at the time of any IM-ITE administration). The specifics of each type of data, their collection, and their individual associations with IM-ITE scores were described previously.¹⁸ Regression coefficients, which represent the expected change in score on the subsequent test for each unit change in score on the previous test, and R-squared, which represents the percent of the variance in each score accounted for by the previous test performance, were

used to describe bivariate associations of subsequent test scores vs previous test scores.

All analyses were conducted by two authors (FSM and SLZ) using Stata 8.2 (StataCorp LP, College Station, TX). Results were considered statistically significant if p-values were less than or equal to 0.05. The study protocol was prospectively approved by the institutional review boards of the Mayo Clinic and the Johns Hopkins Bloomberg School of Public Health. Residents had the opportunity to withhold their data from analysis. All elected to be included in the study.

RESULTS

The association of previous USMLE scores on subsequent USMLE Step scores had correlation coefficients (R-values) ranging from 0.47 to 0.59 ($p < 0.001$). Consequently, the percent of variance (R-squared) accounted for by previous USMLE Step scores ranged from 22% to 35% with the associations gaining in strength with the proximity of one test to another (i.e., USMLE-1 is more predictive of USMLE-2 than of USMLE-3); regression coefficients (R-squared) for USMLE-2 vs. USMLE-1 0.63 (0.35); USMLE-3 vs. USMLE-1 0.43 (0.22); USMLE-3 vs. USMLE-2 0.49 (0.35).

The associations of USMLE Step scores on IM-ITE scores were statistically significant ($p < 0.001$) with the percent variance in IM-ITE scores accounted for by USMLE scores ranging from 17% to 27%: regression coefficients (R-squared) for IM-ITE vs. USMLE-1 0.27 (0.21); IM-ITE vs. USMLE-2 0.29 (0.27); IM-ITE vs. USMLE-3 0.25 (0.17). Previous IM-ITE scores explained 66% of the variance in subsequent IM-ITE scores (regression coefficient 0.75, R-squared 0.66). Assessment of the bivariate associations of all USMLE steps and IM-ITE scores stratified by PGY yielded similar results.

Regression coefficients (95% CI) for the random effects multivariate associations of each USMLE Step with IM-ITE scores were USMLE-1 0.19 (0.12, 0.27), USMLE-2 0.23 (0.17, 0.30), and USMLE-3 0.19 (0.09, 0.29). These associations were adjusted for all 11 variables listed in the methods

Table 2. Scores on United States Medical Licensing Examination (USMLE) Steps 1, 2, 3 and Internal Medicine In-Training Examination (IM-ITE) for Total Cohort

	Number of residents	Number of test administrations	Mean (SD)	Range
USMLE-1*	195	195	230 (15.3)	187 to 266
USMLE-2*	186	186	232 (16.5)	191 to 267
USMLE-3*	89	89	222 (13.8)	182 to 253
IM-ITE†	195	421	65.5% (8.97)	38% to 88%

*USMLE scores are reported as a standardized scaled score with minimum passing score of 180

†IM-ITE scores are reported as total percent items correct. Residents are encouraged to take the IM-ITE each year of internal medicine residency, thus the number of test administrations is greater than the number of residents

DISCUSSION

Investigators who examine the merits of educational initiatives or interventions are tempted to utilize the performance on standardized examinations as an outcome. The IM-ITE is one of the largest standardized post-graduate in-training exams in use. This is the first published study to assess IM-ITE score associations with all three USMLE Step scores appropriately adjusted for previous IM-ITE scores and multiple known and hypothesized predictors of IM-ITE outcome. Our results emphasize the need to take into consideration prior performance when using test scores as an educational outcome.

The percent variance in IM-ITE scores accounted for by USMLE Step scores ranges from 17% to 27%. However, when adjusting for multiple predictors known or hypothesized to affect IM-ITE scores, the regression coefficients for the individual steps of the USMLE on the IM-ITE are quite similar with broadly overlapping confidence intervals. This implies that for multivariate analyses using IM-ITE score as the outcome variable, the use of a single USMLE Step score (e.g., USMLE Step 1) is sufficient to account for the effect of all USMLE Steps on IM-ITE scores. This is important since USMLE Step 1 is almost universally available to residency programs at the time of medical student application to residency, while the other USMLE Steps are not.

As expected, the association of previous IM-ITE score on subsequent IM-ITE score is strong with previous IM-ITE scores accounting for 66% of the variance in subsequent scores. Except for our previous work,¹⁸ we were unable to identify other studies that accounted for prior test performance on subsequent performance.^{14–17,20,23–30} Moving forward, studies that use ITE scores as an outcome should adjust for prior performance on longitudinal repeated measures with appropriate modeling, e.g., random effects methods, as used in this study.^{18,21,22}

Similar to other reports,^{13,31} the association of previous USMLE Step scores on subsequent USMLE Step scores in our study was strong (R ranged 0.47 to 0.59), and the percent of variance (R-squared) accounted for by previous USMLE Step scores ranged from 22% to 35%. Thus, for educational studies that use USMLE Step scores as an outcome, it is important to account for previous scores. However, factors other than previous USMLE Step scores are influencing the variance in these subsequent scores, and these associations should be explored further. The stated purposes of USMLE Steps 1, 2, and 3 are to assess skills and knowledge in the basic sciences, supervised medical care, and independent medical practice, respectively.⁹ These different emphases may account for some of the differences in variances between the individual Steps of the USMLE.

The association of previous test scores on USMLE performance has also been demonstrated. The associations between Medical College Admission Test (MCAT) scores and USMLE Step scores have correlation coefficients of 0.61, 0.49, and 0.49 for USMLE Steps 1, 2, and 3, respectively.³² Donnon et al. recently published a meta-analysis of the predictive ability of the MCAT showing similar results related to USMLE Steps.³³ The association between individual USMLE Steps was not reported in either study. It is notable that the correlation coefficients we report for the effect of previous USMLE Step scores on subsequent USMLE Step scores is similar to those previously reported for MCAT scores. Perhaps this is due to the similar nature of these multiple-choice high-stakes examina-

tions in general or to the similarity of their content in particular.

In addition to the association of MCAT with USMLE scores, the association between IM-ITE performance and subsequent ABIM-CE passage rates has been demonstrated, albeit using much smaller sample sizes.^{4–8} Our study, viewed in context of these previous studies, helps to establish the continuum of association between MCAT, USMLE Steps, IM-ITE, and ABIM-CE.

There are limitations to this study. It is a retrospective cohort analysis from a single institution. This institution has a very high caliber of trainee as noted by the high mean scores on each examination (Table 2). However, the range of scores on each examination span the gamut from very low to very high, so inferences based on this study for scores within these ranges would be appropriate.

USMLE Step scores are often used by residency programs in their admissions criteria. Knowing how the scores of the individual USMLE Steps relate to one another and to the more specific assessment of internal medicine knowledge as determined by the IM-ITE is important to educators. Given our results, we would encourage the developers of the USMLE and the IM-ITE, as well as the ABIM-CE and Re-certifying Examination, to assess the associations of these scores using the entire cohort of learners who take these tests annually. This would further our understanding of inter-test relationships of medical knowledge along the continuum of medical training and allow educators to better interpret factors that may be influencing current performance in relation to prior test scores. Until this is done, analyses such as those presented in our study are important for proper assessment of these tests in educational outcome studies.

Acknowledgments: There was no external funding source for this study. This work was presented at the 13th Ottawa International Conference on Clinical Competence, Melbourne Australia, 3–8 March 2008.

Conflict of interest statement: None disclosed.

Corresponding Author: Furman S. McDonald, M.D., M.P.H.; Residency Program Director and Associate Chair, Department of Medicine, Assistant Professor of Medicine, Mayo Clinic College of Medicine, 200 1st Street SW, Rochester, MN 55905, USA (e-mail: mcdonald.furman@mayo.edu).

REFERENCES

1. ACGME Outcome Project, ACGME General Competencies Vers. 1.3 (9.28.99) Available at: http://www.acgme.org/outcome/comp/comp_full.asp#2. Accessed April 7, 2008.
2. Garibaldi RA, Subhiyah R, Moore ME, Waxman H. The In-Training Examination in Internal Medicine: an analysis of resident performance over time. *Ann Intern Med.* 2002;137(6):505–10.
3. Internal Medicine In-Training Examination®. Available at: http://www.acponline.org/education_recertification/education/in_training/ Accessed April 7, 2008.
4. Grosman RS, Fincher RM, Layne RD, Seelig CB, Berkowitz LR, Levine MA. Validity of the in-training examination for predicting American Board of Internal Medicine certifying examination scores. *J Gen Intern Med.* 1992;7(1):63–7.
5. Cantwell JD. The Mendoza Line and in-training examination scores. *Ann Intern Med.* 1993;119(6):541.
6. Rollins LK, Martindale JR, Edmond M, Manser T, Scheld WM. Predicting pass rates on the American Board of Internal Medicine certifying examination. *J Gen Intern Med.* 1998;13(6):414–6.

7. **Waxman H, Braunstein G, Dantzker D, et al.** Performance on the internal medicine second-year residency in-training examination predicts the outcome of the ABIM certifying examination. *J Gen Intern Med.* 1994;9(12):692-4.
8. **Babbott SF, Beasley BW, Hinchey KT, Blotzer JW, Holmboe ES.** The predictive validity of the internal medicine in-training examination. *Am J Med.* Aug 2007;120(8):735-40.
9. United States Medical Licensing Examination. Available at: <http://www.usmle.org/>. Accessed April 7, 2008.
10. **Pohl CA, Robeson MR, Veloski J.** USMLE Step 2 performance and test administration date in the fourth year of medical school. *Acad Med.* Oct 2004;79(10 Suppl):S49-51.
11. **Melnick DE, Dillon GF, Swanson DB.** Medical licensing examinations in the United States. *J Dent Educ.* May 2002;66(5):595-9. discussion 610-591.
12. **O'Donnell MJ, Obenshain SS, Erdmann JB.** Background essential to the proper use of results of step 1 and step 2 of the USMLE. *Acad Med.* Oct 1993;68(10):734-9.
13. **Andriole DA, Jeffe DB, Hageman HL, Whelan AJ.** What predicts USMLE Step 3 performance? *Acad Med.* 2005;80(10 Suppl):S21-4. Oct.
14. **Cacamese SM, Eubank KJ, Hebert RS, Wright SM.** Conference attendance and performance on the in-training examination in internal medicine. *Med Teach.* 2004;26(7):640-4.
15. **Dupras DM, Li JT.** Use of an objective structured clinical examination to determine clinical competence. *Acad Med.* 1995;70(11):1029-34.
16. **Ellis DD, Kiser WR, Blount W.** Is interruption in residency training associated with a change in in-training examination scores? *Fam Med.* 1997;29(3):184-6.
17. **Hawkins RE, Sumption KF, Gaglione MM, Holmboe ES.** The in-training examination in internal medicine: resident perceptions and lack of correlation between resident scores and faculty predictions of resident performance. *Am J Med.* 1999;106(2):206-10.
18. **McDonald FS, Zeger SL, Kolars JC.** Factors associated with medical knowledge acquisition during internal medicine residency. *J Gen Intern Med.* 2007;22(7):962-8. Jul.
19. **Ogunyemi D, Taylor-Harris de S.** Factors that correlate with the U.S. Medical Licensure Examination Step-2 scores in a diverse medical student population. *J Natl Med Assoc.* 2005;97(9):1258-62. Sep.
20. **Shokar GS.** The effects of an educational intervention for "at-risk" residents to improve their scores on the In-training Exam. *Fam Med.* 2003;35(6):414-7.
21. **Liang KY, Zeger SL.** Regression analysis for correlated data. *Annu Rev Public Health.* 1993;14:43-68.
22. **Diggle P, Heagerty P, Liang K, Zeger S.** Analysis of Longitudinal Data. 2nd edition. Oxford, England: Oxford University Press; 2002.
23. **Pofahl WE, Swanson MS, Cox SS, et al.** Performance standards improve American Board of Surgery In-Training Examination scores. *Curr Surg.* 2002;59(22):20-2. Mar-Apr.
24. **Godellas CV, Huang R.** Factors affecting performance on the American Board of Surgery in-training examination. *Am J Surg.* 2001;181(4):294-6.
25. **Bull DA, Stringham JC, Karwande SV, Neumayer LA.** Effect of a resident self-study and presentation program on performance on the thoracic surgery in-training examination. *Am J Surg.* 2001;181(2):142-4.
26. **Godellas CV, Hauge LS, Huang R.** Factors affecting improvement on the American Board of Surgery In-Training Exam (ABSITE). *J Surg Res.* 2000;91(1):1-4.
27. **Madlon-Kay DJ.** When family practice residents are parents: effect on in-training examination scores. *Fam Pract Res J.* 1989;9(1):53-6.
28. **Pollak R, Baker RJ.** The acquisition of factual knowledge and the role of the didactic conference in a surgical residency program. *Am Surg.* 1988;54(9):531-4.
29. **Black KP, Abzug JM, Chinchilli VM.** Orthopaedic in-training examination scores: a correlation with USMLE results. *J Bone Joint Surg Am.* 2006;88(3):671-6. Mar.
30. **Carmichael KD, Westmoreland JB, Thomas JA, Patterson RM.** Relation of residency selection factors to subsequent orthopaedic in-training examination performance. *South Med J.* 2005;98(5):528-32. May.
31. **Sawhill AJ, Dillon GF, Ripkey DR, Hawkins RE, Swanson DB.** The impact of postgraduate training and timing on USMLE Step 3 performance. *Acad Med.* Oct 2003;78(10 Suppl):S10-12.
32. **Julian ER.** Validity of the Medical College Admission Test for predicting medical school performance. *Acad Med.* Oct 2005;80(10):910-7.
33. **Donnon T, Paolucci EO, Violato C.** The predictive validity of the MCAT for medical school performance and medical board licensing examinations: a meta-analysis of the published research. *Acad Med.* Jan 2007;82(1):100-6.