APM Perspectives

The Association of Professors of Medicine (APM) is the national organization of departments of internal medicine at the US medical schools and numerous affiliated teaching hospitals as represented by chairs and appointed leaders. As the official sponsor of The American Journal of Medicine, the association invites authors to publish commentaries on issues concerning academic internal medicine.

For the latest information about departments of internal medicine, please visit APM's website at www.im.org/APM.

A Commentary on the Use of the Internal Medicine In-Training Examination

Helen Wang,[a] Rebecca Nugent, PhD,[b] Connie Nugent, MLS,[a] Kenneth Nugent, MD,[a] Michael Phy, DO[a]

[a]Department of Internal Medicine, Texas Tech University Health Sciences Center, Lubbock; [b]Department of Statistics, Carnegie Mellon University, Pittsburgh, Penn.

PERSPECTIVES VIEWPOINTS

The Internal Medicine In-Training Examination (IM-ITE) was developed by the American College of Physicians (ACP), the Association of Program Directors in Internal Medicine (APDIM), and the Association of Professors of Medicine (APM).[1] The current test has 340 questions in 11 content areas and a 7-hour time limit, and is administered annually in the fall. The test is designed to evaluate the knowledge base of postgraduate year (PGY)-2 residents; however, most internal medicine residency programs administer the test to residents at all levels. In 2007, more than 18,000 internal medicine residents took the ITE. This test allows residents to compare their medical knowledge with other residents at the same level of training and to enable individual programs to track residents' progress and evaluate and consider curriculum changes. The IM-ITE is a "low-stakes" test, and information on the ACP website suggests residents should not study for this test.[1] We reviewed the literature available on the IM-ITE to determine which factors (if any) correlate with test scores, its utility in predicting a pass on the American Board of Medicine Certifying Examination (ABIMCE), and its usefulness in program development and change.

MATERIALS AND METHODS

We searched the PubMed database using the following MESH terms: "internal medicine," "internal medicine/ education," "internship and residency," and "educational measurement." We used "examination" and "in-training" as text words. Articles also were identified by using the related articles' function on PubMed and by carefully searching the reference lists in identified articles. To analyze the utility of the IM-ITE to predict the results on the ABIMCE, we used sensitivity and specificity. The sensitivity of the IM-ITE is the proportion of residents who passed the ABIMCE who were predicted to pass the IM-ITE. Correspondingly, a false negative is a resident who is predicted to fail the ABIMCE on the basis of his/her IM-ITE results but passes. The specificity is the proportion of residents who failed the ABIMCE who were predicted to fail on the basis of their IM-ITE results. A false positive then is a resident who is predicted to pass the ABIMCE on the basis of his/her IM-ITE results but fails. Both sensitivity and specificity have a maximum value of 1 with higher values indicating higher utility and fewer false positives/negatives. We used the Meta-DiSc software[2] to combine the sensitivities and specificities of individual studies to estimate the overall pooled sensitivity and specificity for the IM-ITE in predicting the results on the ABIMCE.

Test Result Trends

Garibaldi et al[3] analyzed the results from the first 12 years of test administration (1988-2000, total of 13 tests) to internal medicine residents and found that

more than 80% of residents take this test on an annual basis. Test scores increase by approximately 5% with each year of training. There was a significant improvement in the 1995 scores when the time for the test was increased from 6 to 7 hours. Since 1995, international medical graduates have consistently scored higher than graduates from US medical schools, which might reflect the benefit of increased test time for physicians for whom English is a second language. In addition, international medical graduates might have a stronger medical school curricula and more postgraduate experience. PGY-3 residents from US medical schools tested in 1996 to 1998 did not have the expected increase in the average score when compared with the increase observed between PGY-1 and PGY-2 years. Potential explanations for this result included complacency and poor effort by US graduates, excessive moonlighting by US graduates, and other unknown factors. The ACP and associated organizations should update this information annually to track the medical knowledge of current internal medicine residents.

> ### PERSPECTIVES VIEWPOINTS
>
> - Only 1 study suggests that attendance at core curriculum conferences increases ITE scores.
>
> - ITE scores do predict success on the ABIM certifying examination but their utility is limited by frequent false positive and false negative results.
>
> - ITE results have not stimulated formal studies of IM residency curriculum designed to increase scores and improve curriculum management.
>
> - The ITE results provide a rich data base and an opportunity for more complex analysis of knowledge structures in and between content areas.

Test Result Correlates

McDonald et al[4] reported that increased medical knowledge acquisition as measured by IM-ITE performance is associated with regular attendance at educational conferences and use of an electronic resource. Their study involved 195 residents at Mayo Clinic who took the IM-ITE a total of 421 times in 2002 and 2003. These residents attended 4 required residency conferences each week on morbidity and mortality, grand rounds, and 2 core curriculum conferences. Attendance was recorded with an electronic card swipe. The electronic resource used by the residents was UpToDate, which was available to all residents. The IM-ITE percent correct score increased 3.9% per 100 conferences attended and 3.7% per 100 hours spent using UpToDate. These authors also analyzed the correlation between knowledge acquisition measured by IM-ITE scores and specific conference attendance with correction for other factors, including prior medical education variables, such as USLME Step 1 scores, and demographic variables, such as age. A significant correlation was shown between attendance at the core curriculum

series and IM-ITE scores ($P = .04$), but not between attendance at morbidity and mortality series and IM-ITE results ($P = .97$ by multivariate random effects model) or between attendance at the grand rounds series and IM-ITE results ($P = .61$). The overall correlation between total conference attendance and IM-ITE scores after correction for multiple factors, including United States Medical Licensing Examination Step 1 scores, was statistically significant ($r^2 = 0.159$, $P < .001$) but relatively low.[5] In the Mayo Clinic program, core curriculum conferences cover topics developed from a resident needs assessment and are arranged by chief residents to address basic topics in a recurring 2-year cycle.[4] This approach to conference organization might increase the education value because these conferences are focused on the training of general internists.

Cacamese et al[6] found that attendance at conferences did not influence residents' performances on the IM-ITE. Their study included only 19 residents who took the IM-ITE for the first time. This analysis was based on 165 conferences, including 126 at which attendance was taken. After receiving the IM-ITE results, residents filled out a 3-page survey with 15 questions about their study habits and rotations and their opinions about the usefulness of noon conferences using a 5-point Likert scale. The majority of residents believed that their scores, highest or lowest, reflected the amount of clinical exposure they had in the specific subspecialty. However, test scores did not support this conclusion because participation in a subspecialty rotation did not influence either the scores ($r = -0.285$ to $+0.565$, all P values $>.05$) or the conference attendance ($r = -0.168$ to $+0.508$, all $P > .05$). Authors from both residency programs concluded that interactive conferences had the greatest impact on resident learning. These conclusions are based on limited information from 2 residency programs and are necessarily somewhat speculative. More studies should measure the effects of the variables that influence ITE results.

Test Utility

Several studies have evaluated the performance characteristics of the IM-ITE in the prediction of passing the ABIMCE (Table).[7-10] These studies included 949 residents (with a range of 109-398 residents per program) with a pass rate of 66% to 77.5% for the ABIMCE. Using percentiles from 20% to 35% on the ITE resulted in sensitivities that ranged from 0.77 to

Table   Summary of Studies Examining the Utility of the Internal Medicine In-Training Examination in Predicting American Board of Internal Medicine Certifying Examination Results

Study	Training Year	n	ITE Dates	ABIM Pass Rate	ITE Percentile	ITE Utility Sensitivity	Specificity
Grossman et al[8]	PGY 2-3	109	1988, 1989	77.5%	35%	0.95	0.67
Waxman et al[10]	PGY 2	223	1988, 1989, 1990	74.9%	35%	0.80	0.68
Rollins et al[9]	PGY 2	155	1992, 1993, 1994	69.0%	NA	0.88	0.50
	PGY 2	64	1992, 1993, 1994	66.0%	NA	0.77	0.68
Babbott et al[7]	PGY 1	128	1998, 1999, 2000, 2001, 2002	72.0%	24%	0.81	0.82
	PGY 2	136	1998, 1999, 2000, 2001, 2002	72.0%	23%	0.83	0.84
	PGY 3	134	1998, 1999, 2000, 2001, 2002	72.0%	20%	0.84	0.82
Overall		949		Range of 66.0%-77.5%	Range of 20%-35%	0.83[a] (0.80-0.85)	0.75[a] (0.69-0.80)

ABIM = American Board of Internal Medicine; ITE = In-Training Examination; NA = not available; PGY = postgraduate year.
[a]Pooled sensitivity/specificity measures with 95% CI.

0.95 and specificities that ranged from 0.50 to 0.84 for passing or failing the ABIMCE.[2] We calculated a pooled sensitivity of 0.83 (95% confidence interval, 0.80-0.85) and a pooled specificity of 0.75 (95% confidence interval, 0.69-0.80). The IM-ITE does predict success on the ABIMCE but has both false-positive and false-negative results. A pooled sensitivity of 0.83 indicates that residents frequently pass the ABIMCE with poor scores on the IM-ITE. This behavior is probably explained by differences in the rate of acquisition of medical knowledge and skills needed to analyze clinical scenarios during residency training, the quality of effort during preparation for the ABIMCE, and the educational environment in internal medicine programs.

Grossman et al[8] found a significant correlation between the composite percentile score on the ABIMCE and IM-ITE ($r^2 = 0.58$ for PGY-2 residents, 0.68 for PGY-3 residents, $P < .001$ for both correlations) and a weaker and variably significant correlation ($r^2 = 0.041$-0.319 with P values ranging from <0.01 to >0.05) between the scores on the subspecialty components of these 2 tests. The authors also determined that the ABIMCE pass rate correlated positively with the IM-ITE result and negatively with age.[11] Between 2003 and 2007, the pass rate on the ABIMCE was 91% to 94%. The pass rate is based on an absolute standard set by the writing committee using a modification of the Angoff method. It is criterion-based and not normative-based.[12] This pass rate is higher than the pass rates in Table; this change presumably reflects a change in standard. Whether the current IM-ITE scores have the same diagnostic utility for predicting ABIMCE results is unknown and needs to be calculated using current resident cohorts.

Program Changes Based on Test Results
The IM-ITE results offer internal medicine programs the opportunity to critically evaluate their programs, which most program directors do annually. However, to our knowledge, no studies report changes in curricula or didactic methods based on IM-ITE results. Several studies have reported relatively low conference attendance by residents with a decline during the academic year. Cacamese et al[6] suggested that IM-ITE results do not correlate with conference attendance or recent rotations on subspecialties that might influence the development of medical knowledge in particular areas.[11,13] Therefore, the best approach to program change based on IM-ITE scores is unclear and changes in the residency curriculum may not have as much impact as might be expected. de Virgilio et al[14] reported that weekly assigned reading and multiple choice tests on the reading significantly improved IM-ITE scores in a surgery residency. This change requires significant effort by the program director or a designate. The more likely strategy used by program directors is to select, when possible, residents with high scores on the United States Medical Licensing Examination Steps 1 and 2 and depend on resident motivation and ability to drive ABIMCE pass rates.

DISCUSSION
IM-ITE results comprise a large database that reflects the working knowledge of internal medicine residents. This information provides an opportunity to measure trends and effectiveness in both medical education and residency training, particularly if a core set of questions is used annually to allow valid comparisons between resident cohorts. These results can generate multiple comparisons, including differences between US medi-

cal graduates and international medical graduates and differences among US medical schools. These results allow an individual program to evaluate resident performance and to evaluate its curriculum and didactic methods. The IM-ITE also provides a reasonable predictor of outcomes on the ABIMCE. However, given the current pass rate ($\geq$90%), most residents will pass with the usual preparation and educational curricula available in most programs and would not need information about medical knowledge from IM-ITE. In addition, the false-positive results on the IM-ITE may create complacency in some residents with disastrous results. IM-ITE scores allow individual residents to compare their current status with their peers at the same training level, identify areas of weakness, and evaluate learning methods. However, most residents probably do not use this opportunity, decreasing the utility of IM-ITE. Finally, the conclusions developed from the several studies on IM-ITE results have not been rigorously tested and only reflect the conclusions of a few program directors using a small data set. No internal medicine program has reported studies describing changes in curriculum based on IM-ITE scores.

The value of IM-ITE could be enhanced with careful evaluation by ABIMCE and IM-ITE test-writing committees to determine whether these tests represent core knowledge needed to practice internal medicine. The ABIMCE is reviewed for clinical relevance (self-defined by individual reviewers), and these questions likely reflect core material. However, the correspondence about ideas regarding core knowledge and clinical relevance, if any, between the 2 committees is unknown. In particular, program directors need to know whether the information tested on the IM-ITE represents the information tested on the ABIMCE. These committees might develop a list of core knowledge objectives that could be used to direct reading and organize didactic conferences. The ACP, APDIM, and APM have the opportunity to use these results to evaluate differences in curriculum in various information resources and didactic methods. These studies would require extensive preparation to develop adequate cross-sectional and longitudinal protocols to evaluate program organization and didactics. However, identifying programs with consistent success and analyzing their characteristics seem useful. In addition, these IM-ITE results could be used to evaluate optimal learning types and methods for future internists and to evaluate medical school curricula and subsequent success in internal medicine. It is possible that medical students entering internal medicine residencies should do more work in internal medicine preparation during their fourth year because the core knowledge base is extensive and continuously increasing.

The IM-ITE can provide a rich database that includes results from individual residents, individual programs, training year, calendar year, medical school education, and content area. In addition, test questions could be classified according to Bloom's taxonomy into groups that require understanding, interpretation, and analysis.[15] This information from IM-ITE might allow us to determine hierarchic structures in the content areas and in-training level. These (possibly partial) ordered knowledge structures allow analysts to determine whether there is some optimal ordering for learning specific skills or acquiring knowledge.[16-18] For example, it should be possible to determine whether third-year residents consistently answer questions correctly that require analysis more often than first-year residents. Moreover, it might be possible to determine whether doing well on one particular content area requires prior experience or knowledge in another content area. Finally, it also might be possible to determine whether residents require educational maturity to succeed in a particular area. For example, managing complex patients with rheumatologic disorders might require a certain number of months of experience in general internal medicine and should be scheduled in the third year. This information is available from the IM-ITE and requires the ACP and related organizations to undertake additional analyses.

RECOMMENDATIONS

1. The ACP should update and publish annually the information from ITE results to identify current trends. More sophisticated analysis of the results can provide information about educational structures not apparent in simple summaries of the percentage of correct answers.

2. The ACP, ABIM, and APDIM should reach consensus regarding the core knowledge needed to practice general internal medicine. The test-writing committee should make certain that this knowledge is covered on the IM-ITE, enhancing its value as a practice test for the ABIMCE.

3. Information from PGY-1 score results should be provided to medical schools. Changes in the fourth-year curricula might improve the core knowledge of students entering internal medicine residencies.

4. The ACP, ABIM, APDIM, and Residency Review Committee on Internal Medicine should sponsor studies to evaluate the effect of residency curriculum and organization on IM-ITE scores. This test is the only quantitative end point available to evaluate internal medicine residency training. For example, it is possible that the widespread adoption of the night float systems has reduced the core knowledge of internal medicine residents.

References

1. American College of Physicians. Internal Medicine In-Training Exam. Available at: http://www.acponline.org/education_recertification/education/in_training/. Accessed May 21, 2009.

2. Zamora J, Abraira V, Muriel A, et al. Meta-DiSc: a software for meta-analysis of test accuracy data. *BMC Med Res Methodol*. 2006;6:31.

3. 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;137:505-510.

4. McDonald FS, Zeger SL, Kolars JC. Factors associated with medical knowledge acquisition during internal medicine residency. *J Gen Intern Med*. 2007;22:962-968.

5. McDonald FS, Zeger SL, Kolars JC. Associations of conference attendance with internal medicine in-training examination scores. *Mayo Clin Proc*. 2008;83:449-453.

6. Cacamese SM, Eubank KJ, Hebert RS, Wright SM. Conference attendance and performance on the in-training examination in internal medicine. *Med Teach*. 2004;26:640-644.

7. Babbott SF, Beasley BW, Hinchey KT, et al. The predictive validity of the internal medicine in-training examination. *Am J Med*. 2007;120:735-740.

8. Grossman RS, Fincher RM, Layne RD, et al. Validity of the in-training examination for predicting American Board of Internal Medicine certifying examination scores. *J Gen Intern Med*. 1992;7:63-67.

9. Rollins LK, Martindale JR, Edmond M, et al. Predicting pass rates on the American Board of Internal Medicine certifying examination. *J Gen Intern Med*. 1998;13:414-416.

10. 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;9:692-694.

11. Grossman RS, Murata GH, Fincher RM, et al. Predicting performance on the American Board of Internal Medicine Certifying Examination: The effects of resident preparation and other factors. Crime Study Group. *Acad Med*. 1996;71(10 Suppl):S74-S76.

12. American Board of Internal Medicine. How exams are developed. Available at: http://www.abim.org/about/examInfo/developed.aspx. Accessed May 21, 2009.

13. FitzGerald JD, Wenger NS. Didactic teaching conferences for IM residents: who attends, and is attendance related to medical certifying examination scores? *Acad Med*. 2003;78:84-89.

14. de Virgilio C, Stabile BE, Lewis RJ, Brayack C. Significantly improved American Board of Surgery In-Training Examination scores associated with weekly assigned reading and preparatory examinations. *Arch Surg*. 2003;138:1195-1197.

15. Bloom BS. *Taxonomy of Educational Objectives, Handbook 1: Cognitive Domain*. Reading, MA: Addison Wesley Publishing; 1956.

16. Desmarais M, Gagnon M, Meshkinfam P. Item-based Bayesian student models. In: *Proceedings of the 21st National Conference on Artificial Intelligence*. Menlo Park, CA: 2006.

17. Doignon JP, Falmagne JC. *Knowledge Spaces*. New York, NY: Springer; 1999.

18. Falmagne JC, Koppen M, Villano M, et al. Introduction to knowledge spaces: how to build, test, and search them. *Psychol Rev*. 1990;97:201-224.