

The ASN In-Training Examination and the ABIM Certifying Examination: Time for a New Testing Paradigm

Robert S. Brown

Department of Medicine, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, Massachusetts

In 2009, a new computer-based in-training examination (ITE) was created by the American Society of Nephrology (ASN) in conjunction with the National Board of Medical Examiners to assess the medical knowledge of fellows training in nephrology. The ITE, patterned after the American Board of Internal Medicine certifying examination for nephrology (ABIM-CE), allowed fellows and training program directors (TPDs) to compare their performance with programs in the nation. However, despite using case vignettes to simulate real-life clinical experiences, the nation's graduating fellows answered a mean of only 69% of the questions correctly. Although unpublicized, the ABIM-CE passing grade can be estimated to be approximately 65%, resulting in the current passing rate of 94%. The ABIM states that they are testing 85% synthesis and clinical judgment, rather than recall memory. So are we to conclude from these two exams that the medical knowledge of our graduating nephrology fellows allows them to properly manage only about two-thirds of their patients? Rather, the construction of the test questions and their relevance to clinical practice should be reassessed. The time has come for a new testing paradigm, one that would require candidates passing the ABIM-CE to score 85% to 90% correct answers to more relevant "core" examination questions that assess the actual medical knowledge needed to practice competently in our subspecialty. If the ABIM-CE was revised, the ASN ITE, with no specific passing grade, could then test core and "factual" recall knowledge with feedback of questions to TPDs to assess and help fellows in training.

Clin J Am Soc Nephrol 5: 1513–1517, 2010. doi: 10.2215/CJN.02050310

As a long-term nephrology training program director (TPD), I applaud Mitch Rosner and his colleagues on the American Society of Nephrology (ASN) In-Training Examination (ITE) Committee for developing the first nationwide examination for nephrology fellows in training (1). In conjunction with the National Board of Medical Examiners (NBME), the ASN developed and validated 150 multiple choice questions to serve as "a comprehensive assessment of medical knowledge" possessed by our current fellows. The questions were in 11 broad "content domains chosen to mirror the blueprint of the American Board of Internal Medicine certifying examination for nephrology (ABIM-CE) (Table 1). The questions were written in a case-based vignette format to be compatible with real-life clinical experiences" (1). In April 2009, 301 first-year fellows and 381 second-year and higher fellows (approximately 84% of the estimated 812 nephrology fellows in training in the United States) took this 6-hour computer-based exam.

Results and Analysis of the ITE

Each TPD was given the fellows' individual scores in each content topic and overall, together with data to compare each fellow's scores and the program's scores with national averages. Thus, the first of the five goals of the ASN ITE Com-

mittee (1), "to provide TPDs with information regarding the potential strengths and weaknesses in their individual programs through benchmarking of their program compared with other programs on the basis of scores on the examination," was successfully achieved. The fellow examinees were shown their score within each content area and their overall score, together with a comparison with their national peer group by fellowship year. Table 1 depicts the mean percentage of correct answers and the SD achieved by first-year fellows, second-year and higher fellows, and all fellows in each content area and on the total ITE. Surprisingly, the nation's graduating second-year fellows only achieved a mean score of 69% \pm 7% correct answers, or approximately 95% of graduating fellows scored between 55% and 83% overall. In our program, we offered the ITE to our second-year fellows, all of whom took the exam to prepare for the ABIM subspecialty certifying exam in nephrology and because I, as their TPD, requested them to do so. The average total score of our program's fellows was quite high—in the 88th percentile of the second-year test-takers. However, this translated to a mean total score of only 76.9% correct answers and their average scores in the 11 content areas ranged from 57% to 88% (the range of content area scores for individual fellows varied from 50% in ethics to 100% in mineral metabolism). Neither the fellows nor the TPD were provided the results of individual questions or the topics of individual questions. Without such data, the second goal of the ITE, "to provide feedback to fellows in-training regarding their relative strengths and weaknesses in specific content areas as

Published online ahead of print. Publication date available at www.cjasn.org.

Correspondence: Dr. Robert S. Brown, Nephrology Division, Beth Israel Deaconess Medical Center, 185 Pilgrim Road, Boston, MA 02215. Phone: 617-632-9880; Fax: 617-632-9890; E-mail: rbrown@bidmc.harvard.edu

Table 1. 2009 ASN ITE content areas and mean examinee performance as percent correct scores by fellowship year^a

| Content Area | Questions (%) Devoted to Content Area | First-Year Fellows ^b (n = 301) | Second-Year and Higher Fellows ^{b,c} (n = 381) | All Fellows ^b (n = 682) |
|--|---------------------------------------|---|---|------------------------------------|
| General aspects of chronic kidney disease | 13 | 65 (12) | 71 (11) | 68 (12) |
| Glomerular/vascular | 12 | 57 (12) | 61 (12) | 60 (12) |
| Tubular/interstitial and cystic | 6 | 74 (16) | 81 (16) | 78 (17) |
| Acute renal failure/intensive care unit nephrology | 10 | 61 (11) | 63 (11) | 62 (11) |
| Kidney transplantation | 10 | 52 (14) | 58 (15) | 55 (15) |
| Hypertension | 10 | 65 (13) | 68 (12) | 67 (12) |
| Electrolyte physiology: sodium/water | 9 | 74 (14) | 78 (13) | 76 (13) |
| Electrolyte physiology: acid-base/potassium | 9 | 57 (15) | 61 (14) | 59 (15) |
| Mineral metabolism | 8 | 72 (16) | 80 (13) | 77 (15) |
| Clinical pharmacology | 9 | 70 (13) | 73 (13) | 72 (13) |
| Ethics | 4 | 75 (21) | 76 (20) | 76 (20) |
| Total test | 100 | 64 (8) | 69 (7) | 67 (8) |

^aAdapted from Tables 1 and 2 from Rosner *et al.* (1).

^bValues presented as mean percentage (\pm SD).

^cThis group also includes examinees in their third, fourth, and fifth year of fellowship.

well as to allow them to compare their results against national outcomes," was only partially achieved. The ASN ITE Committee states that "there were several reasons why this information (*i.e.*, the questions and answers) was not provided: (1) the examination is not meant to have programs or trainees study specific topics but is designed to provide an overall assessment of medical knowledge competency; and (2) within each content area the SD can be high because the number of items are few, thus a content-specific score may be misleading" (1). Although these reasons have some validity, I think that the counterargument that the examinees and TPD would learn much more by reviewing the questions and answers together should have greater weight. With relatively few questions in each broad content area, it does not seem reasonable for a fellow to base a conclusion that they are strong or weak in a content area on the basis of answers to as few as two to four unknown questions. Likewise, the third goal of the ITE, "to facilitate identification and discussion of potential weaknesses in education in nephrology training programs nationally," is also hampered by the lack of specific data about the questions. Because of the wide variation among fellows and middling scores in different content areas, it was difficult for TPDs to conclude whether their programs were deficient in a given content area, much less to make assumptions about "weaknesses in education in nephrology training programs nationally." Perhaps the strongest motivation to maintain the secrecy of these test questions is based on the cost of constructing new questions each year, and another rationale may be keeping the subject matter a mystery to avoid "teaching to the test." The achievement of the fourth goal, "to allow fellows to track changes in their medical knowledge between their first and second years of renal fellowship," may be realized next year. But for now the increase in the test scores between first- and second-

year fellows was a surprisingly low 5%, from 64% to 69% (Table 1). Perhaps this will improve further when the same first-year fellows take the test again as second-year fellows in 2010, but improvements on taking a similar test a second time are open to other explanations related to test-taking rather than a more comprehensive knowledge of the subject matter. Did the ITE test achieve its fifth goal, "to allow assessment of medical knowledge competency in trainees"?

If these questions truly represent case formats that are consistent with "real-life clinical experiences," having only half of graduating fellows answering more than 69% correctly seems to offer a disastrous picture of our nephrology training programs or of our fellows' medical knowledge competency. The U.S. public deserves better than specialists who can correctly diagnose, manage, understand, or answer questions about only two-thirds of the "real-life clinical" patient situations in which they are trained. A counterargument might be made that this test situation does not show that the trainees would necessarily be wrong one-third of the time in practice, but it certainly has not documented that they would be correct either. The ASN ITE Committee did not comment on whether these results met or failed to meet their expectations. Only a systematic review of the questions, whether they tested good clinical judgment, memorized trivia, or tricky interpretations, can really assess this. But the test results certainly failed to meet my expectations. I know our program's fellows, and those that took the ITE are capable nephrologists. However, their scores, although well above average with a mean of 76.9%, do not accurately depict their excellent clinical knowledge in these content domains. I strongly suspect that the 50% of senior fellows in the nation scoring below 69% also know much more about clinical patient diagnosis, disease management, and ethics than these abysmal scores seem to indicate. So I must conclude that the data that I cannot evaluate—the questions asked on the ITE—are the cul-

perts here. The ASN ITE Committee did a conscientious job of vetting the questions and answers. First, the NBME identified 13 of the 150 items that met certain criteria (*e.g.*, “fewer than 30% of the examinees responded correctly”), but the NBME excluded only 4 of these 13 items and changed one answer. Second, the ASN ITE Committee conducted a post-test survey completed by 84% of the fellows in which only 73% thought the exam was at the appropriate level of difficulty, but more telling, only 41.7% considered the exam to be very relevant with another 34.9% moderately relevant. My assumption is that nephrology TPDs and the U.S. public have an expectation that our specialists should, at a minimum, be able to manage at least 85% to 90% of “real-life clinical experiences” and almost all should consider the examinations designed to document competency to be very relevant. If that is our expectation, then we must change our testing paradigm.

The ABIM-CE

The ITE was constructed “to mirror the blueprint of the ABIM certifying examination for nephrology,” widely held as the standard of medical knowledge. Of the eight fellows from our program taking the ABIM-CE from 2006 to 2008, despite their overall results being well above the national average, they only correctly answered a mean of 78.7% of the 200 total questions on the ABIM-CE. The close similarity of the scores achieved on the ITE and the ABIM-CE by fellows graduating from our program, although in different years, certainly argues that the ITE has mirrored the ABIM-CE or at least is testing for similar medical knowledge bases. Therefore, it does seem quite likely that most of those fellows scoring above the 20th percentile on this ITE will likely pass the ABIM nephrology boards in the future, similar to the results found for the internal medicine ITE as a valid predictor of performance of medical residents later taking their ABIM certifying examination in general internal medicine (2). But because I am questioning whether the nephrology ITE is an appropriate gauge of medical knowledge needed for clinical care, we should consider whether its blueprint, the ABIM-CE, also satisfies this purpose. The ABIM-CE uses an extensive process to develop its questions and states that “the overwhelming majority of ABIM examination questions use patient-based formats assessing the higher-order cognitive abilities required for clinical decision-making” and only what the examinee “is expected to know without access to medical resources or references, as opposed to knowledge that is appropriate—or even mandatory—to ‘look up’” (3). Similar to the ITE, the questions and answers on the ABIM-CE are not provided to the TPD or the examinees, nor is the passing grade. However, on the basis of the scores and deciles of our test-takers between 2006 and 2008, I estimate the passing score to be approximately 65% correct answers to the 200 questions asked. (Of course, the actual ABIM-CE passing score would be welcome information to TPDs.) As described by the ABIM, “the minimum passing score reflects an absolute standard that is independent of the performance of any group of candidates. This standard has been established by the examination committee and approved by the ABIM Board of Directors” (4). Thus, presumably using a form of the Angoff method to iden-

tify the score that separates the borderline competent from the incompetent test-taker (5), the ABIM panel of judges considered those answering approximately two-thirds of the questions to be competent. The content of the ABIM-CE is said to represent 20% to 35% synthesis, greater than 50% clinical judgment, and less than 15% recall knowledge (6). If this were to be true, we should worry about the synthesis skills and clinical judgment of our country’s nephrology specialists because their knowledge, on the basis of their ABIM-CE results, appears to be inadequate to properly synthesize or care for roughly one-quarter to one-third of the patients that they are likely to see, particularly those seen under the time constraints of current office practice. Rather, I think our certifying examinations, although updated, still represent somewhat elitist holdovers from a past time when they were intended to separate the trained internist from the generalist or the trained nephrologist from the internist by testing recall of less common facts. In an era of easy and rapid access to electronic information that can help the clinician with such data, this type of exam is no longer optimal. And because the ABIM is now passing 91% to 94% of first-time takers of the ABIM-CE in nephrology (7), is it not time to change the testing paradigm to reflect what we really want our graduate fellows to know? Why would we want to give a test on which the passing grade may be set as low as 65% by the ABIM examination committee that considers 91% to 94% of the examinees to be satisfactory? Does it not matter which two-thirds of the questions a passing examinee knows and does not know? Moreover, when scores are that low, almost one-third of the correct answers of a borderline candidate (20% of the questions on the basis of five multiple choices) may be obtained just by guessing. It is time to change!

A New Testing Paradigm

Therefore, to accomplish such a change with broad input from U.S. training programs and at minimal cost, I would suggest the following as a possible solution:

1. The ASN ITE Committee should provide the ITE questions and answers to the TPDs and examinees for review after the exam in 2010. Only through review of the questions and answers can trainees and TPDs evaluate what the fellows know and need to learn from their training programs.
2. Each TPD should be asked to submit approximately three questions to the ASN ITE Committee for the following year’s exam. Certainly, this will solve the requirement to make new questions each year and perhaps make the questions more relevant. Moreover, because all data will be provided to the trainees and TPDs and because there is no passing grade for the ITE, there would be less need to vet the answers. The TPDs might receive different domain assignments to avoid too much overlap, only leaving the ASN ITE Committee the task of selecting the best questions.
3. The ABIM Nephrology Subspecialty Board should produce a certifying exam that primarily tests synthesis and clinical judgment in nephrologic patients. The questions should more accurately reflect the skill sets needed by our fellows to give an adequate standard of medical care to patients, not

perpetuate the type of exam that tests uncommon conditions or lesser known facts, in part to obtain a wide distribution of scores. After all, this is essentially a pass-fail exam to certify specialists, not to offer grades that rank who are our smartest nephrology trainees. The expectation would be that a well trained, board-certified nephrologist should be able to correctly answer at least 90% of questions relevant to their specialty practice, and graduating fellows that pass the ABIM-CE should be able to correctly answer at least 85% to 90% of questions. A criticism might be made that such a change just creates an “easier” test with a higher passing score, not a “better” test. But by being more relevant, the new test will result in an ABIM panel of judges requiring a higher passing score to demonstrate borderline competence. Such a test should better document that passing candidates know the “right stuff” with a lower percentage of correct answers based on guessing.

4. One way to construct such an examination for the ABIM-CE might be to separate the questions into two basic categories: a “core” category of questions that all certified nephrologists should know and a “factual” category that could be based on more recall memory. The core category could contain approximately 75% to 80% of the test questions, and the examinee would be required to get at least 85% to 90% correct or would fail the exam. The factual category would serve to test knowledge of the major nephrologic conditions but would play a lesser role in passing the exam. Overall, the percentage of examinees passing the exam might still end up at 91% to 94% (although the actual percentage passing could be lower or greater), but at least the TPDs and the public would be better assured that these certified subspecialists have the knowledge required to practice competently as nephrologists. The risk that this change may worsen the quality of care provided by our certified specialists should be quite low because the evidence that the current certification process is correlated with quality outcomes is, at best, marginal and controversial now (8–11). Moreover, the sentiment among physicians to enroll in the current ABIM maintenance of certification program is strongly negative, with only 37% of 2512 respondents overall, and 23% and 39%, respectively, of board-certified physicians with and without “grandfather” status recommending enrollment (12).
5. If the ABIM changes its test to this new format, the ASN ITE Committee could continue to produce an exam that closely mirrors the ABIM-CE, if desired, but it need not do so in the new paradigm. Because the ASN ITE is not a pass-fail exam and has no set passing grade, I would suggest a two-part exam weighted quite differently. A smaller core component, perhaps 25% to 35% of the questions, might mirror the new ABIM-CE, testing primarily synthesis and clinical judgment to prepare fellows for the ABIM-CE, whereas a larger factual component would help trainees know the areas of nephrology in which they need to concentrate their learning. Of course, this assumes that the fellows and TPDs are given the questions and answers to review after this test. The construction of a new ASN ITE each year should obviate any incentive to “teach to the test.”

In summary, the ASN ITE Committee has produced an exam for nephrology fellows in-training that mimics its prototype, the ABIM-CE. The ITE allows fellows to assess weaknesses and strengths in their factual knowledge, to compare their performance with fellows on a national level, and to prepare for the ABIM-CE. The ITE allows TPDs to compare how their fellows performed with other programs and perhaps to detect content areas that might be improved. However, the failure to provide the TPDs with the questions and answers seriously undermines the ability of the examinee and the TPD to assess their knowledge gaps and reduces the value of the test to provide meaningful improvements to many programs. More seriously, if the ITE serves “to facilitate identification and discussion of potential weaknesses in education in nephrology training programs nationally,” what are we to conclude from learning that approximately half of our graduating fellows scored less than 69% correct answers? It is that this ASN ITE and the ABIM-CE upon which it is patterned, both designed “to allow assessment of medical knowledge competency in trainees,” fail to tell us what we really want to know—whether they are competent to practice satisfactorily in the subspecialty of nephrology. It is time for a new testing paradigm that revises the pass-fail ABIM-CE to a more relevant core exam, and the ASN ITE would no longer need to mirror the ABIM-CE and could test core and factual recall-based knowledge of nephrology.

Disclosures

None.

References

1. Rosner MH, Berns JS, Parker M, Tolwani A, Bailey J, DiGiovanni S, Lederer E, Norby S, Plumb TJ, Qian Q, Yeun J, Hawley JL, Owens S (The ASN In-Training Examination Committee): Development, implementation, and results of the ASN In-Training Examination for fellows. *Clin J Am Soc Nephrol* 5: 328–334, 2010
2. Babbott SF, Beasley BW, Hinchey KT, Blotzer JW, Holmboe ES: The predictive validity of the internal medicine in-training examination. *Am J Med* 120: 735–740, 2007
3. American Board of Internal Medicine: How exams are developed. Available online at <http://www.abim.org/about/examInfo/developed.aspx>. Accessed March 2, 2010
4. American Board of Internal Medicine Certification: Dates, blueprints and scoring. Available online at <http://www.abim.org/exam/cert/neph.aspx#scoring>. Accessed April 5, 2010
5. Verheggen MM, Muijtjens AM, Van Os J, Schuwirth LW: Is an Angoff standard an indication of minimal competence of examinees or of judges? *Adv Health Sci Educ Theory Pract* 13: 203–211, 2008
6. American Board of Internal Medicine Board Certification: A path to quality care. Available online at <http://www.abim.org/pdf/cert-related/ABIM-Brochure-cx.pdf>. Accessed March 2, 2010
7. American Board of Internal Medicine: First-time taker pass rates—Initial certification. Available online at <http://www.abim.org/pdf/pass-rates/cert.pdf>. Accessed April 8, 2010

8. Chen J, Rathore SS, Wang Y, Radford MJ, Krumholz HM: Physician board certification and the care and outcomes of elderly patients with acute myocardial infarction. *J Gen Intern Med* 21: 238–244, 2006
9. Norcini JJ, Lipner RS, Kimball HR: Certifying examination performance and patient outcomes following acute myocardial infarction. *Med Educ* 36: 853–859, 2002
10. Landon BE: What do certification examinations tell us about quality? *Arch Intern Med* 168: 1365–1367, 2008
11. Levinson W, King TE Jr, Goldman L, Goroll AH, Kessler B: American Board of Internal Medicine maintenance of certification program. *N Engl J Med* 362: 948–952, 2010
12. Kritek PA, Drazen JM: Clinical decisions: American Board of Internal Medicine maintenance of certification program—Polling results. *N Engl J Med* 362: e54, 2010

See related editorial, “The ASN In-Training Examination Needs More Time, Not a New Paradigm,” on pages 1363–1365.