

unheeded by legislators and policymakers. In the generation since, several of the issues it identified—including the limitations of fee-for-service medicine and the need to buttress the primary care workforce—have grown more acute.

NASEM's study acknowledges that urgency with a 5-pronged implementation plan to make high-quality primary care available and accessible nationwide. Specifically, it calls for policies that

- Pay for primary care teams to care for people, not doctors to deliver services
- Ensure that high-quality primary care is available to every individual and family in every community
- Train primary care teams where people live and work
- Design information technology that serves the patient, the family, and the interprofessional care team
- Ensure that high-quality primary care is implemented in the United States

The report's findings and recommendations support the Academy's position that the country's fee-for-service health care design promotes misaligned incentives and prizes "sick care" at the expense of population wellness. This dangerous gap was exposed and exacerbated by the COVID-19 pandemic.

AAFP EVP and CEO Shawn Martin said in a statement that coincided with publication of the report, "We look forward to working with policymakers, payers and our other partners in primary care to make the study recommendations a reality—the health of our nation depends on it."

Academy President Ada Stewart, MD, of Columbia, South Carolina, added: "The NASEM report clearly spells out the case for increased investment in our primary care system and ensuring everyone in our country has access to high-quality primary care, something the AAFP has long advocated for. The COVID-19 pandemic further exposed flaws in our current health care system, including those related to many years of underinvestment in primary care."

*AAFP News*



From the American  
Board of Family Medicine

*Ann Fam Med* 2021;19:377-379. <https://doi.org/10.1370/afm.2726>.

## THE EVOLUTION OF KNOWLEDGE ASSESSMENT: ABFM'S STRATEGY GOING FORWARD

Knowledge is a foundation of the public trust in physicians. Knowledge drives diagnosis, treatment, and the shared decision making central to health care. Assessment of medical knowledge has thus been a cornerstone of Board Certification since 1914, when the American Board of Ophthalmology developed the first Board examination. As ABFM rethinks its certification portfolio, it is appropriate to revisit the scientific rationale for assessment of knowledge. What follows frames the key questions and evidence that drive ABFM policy and describes next steps.

### Do Physicians Know What They Know?

It is common for physicians to be confident about what they know. Unfortunately, however, the accuracy of self-assessment of knowledge is uneven and often poor. In the 1990s, Kruger and Dunning conducted a series of experiments with undergraduate students examining their ability to self-assess against objective criteria "in [1 of 4] domain[s] in which knowledge, wisdom, or savvy was crucial: humor, logical reasoning, and English grammar."<sup>1</sup> In each test, there was a relatively narrow range of self-perceived ability, but a much wider range was seen in actual test scores. In each case, the highest objective performers somewhat underestimated their ability and performance. Importantly, however, the lowest performers substantially overestimated their abilities. Subsequently, there has been substantial literature documenting that the Dunning-Kruger effect is pervasive across different professions,<sup>2</sup> including medicine.<sup>3</sup>

Further complicating self-assessment of clinical knowledge is the challenge of keeping up to date. Modern health care is dynamic, with ongoing and important changes in practice standards. There is good evidence that more experienced physicians may be less likely to apply up to date practice guidelines.<sup>4</sup> This is particularly challenging in a generalist specialty like family medicine. In recent years, for example, consider recent changes in evidence and practice in areas including COVID diagnosis and management, a new generation of effective agents for diabetes, point-of-care

ultrasound and new guidelines for depression screening and overuse of x-rays for back pain.

The major goal of ABFM's major knowledge assessments is thus to provide objective, independent assessment of the knowledge necessary to be board-certified. This is true for both the 1-day Family Medicine Certification Examination and the Family Medicine Certification Longitudinal Assessment (FMCLA).

### Is Assessment of Clinical Knowledge Valid and Fair?

Concerns about bias in standardized testing for undergraduate admissions have led some to question all multiple-choice testing. ABFM, however, believes that assessment of clinical knowledge can be valid and fair. Psychometrics emerged in the 1880s<sup>5</sup>; since then, many procedures have been developed for assessing specialty-specific clinical knowledge.<sup>6,7</sup> Procedures for setting standards<sup>8</sup> and for detecting possible<sup>9,10</sup> biased questions for or against specific groups of examinees have been developed. Standards for testing organizations<sup>11,12</sup> have been developed to promote best testing practices. ABFM's strategy for certification, quality control and revalidation have been anchored in these foundations.

Some clinicians believe that it is easy to write multiple choice questions. Like meta-analysis, however, writing clinical multiple-choice questions is easy to do but hard to do well. Substantial effort goes into the development of questions used for ABFM certification. Practicing clinicians write questions, experienced editors work on them, and 2 different committees of practicing clinicians review them. ABFM pre-tests all questions on large samples of examinees, and, since 2013, has regularly used differential item functioning analysis<sup>10</sup> to review all examination questions for evidence of racial, ethnic, or gender bias. Although a few questions have been removed as the result of this ongoing review, there seems to be no substantial systematic question-level bias: indeed, our questions slightly favor underrepresented minorities.<sup>13</sup> We are committed to continuing this review process and will extend it to rurality.

An important advantage of standardized testing is fairness compared with oral examinations and other alternatives. Setting a passing standard that is uniform for all examinees is crucial. The ABFM reviews the passing standard once every 3 years. To help set the standard, ABFM collects extensive data from a pool of participants—much larger than most certification boards. Although setting a standard includes many different analyses, it is ultimately a policy decision. The ABFM Board of Directors reviews the results from several standard-setting procedures<sup>8</sup> as well as information about the history of the passing standard over

time and predictions for future pass rates. They then decide to either retain the existing standard or increase or decrease it by a specific amount.

### Can Board Certification Support Learning and Keeping Up to Date?

Although the major purpose of ABFM certification examinations is to assess knowledge on behalf of the public, there has been also substantial recent interest in using the certification process to support learning and keeping up to date. There is substantial evidence that knowledge decays over time. A "forgetting curve" was first empirically described by Ebbinghaus.<sup>14</sup> Studies have since confirmed that, without reinforcement, much newly learned information is quickly forgotten, with further forgetting occurring more slowly over time.<sup>15</sup> How and at what level of granularity this applies to clinical knowledge is less clear. Other knowledge, experiences, and cognitive processes as well as normal aging can interfere with information retrieval.<sup>16</sup>

Assessment can drive learning. Ample evidence published over the past 20 years shows the knowledge retrieval that occurs with testing is more effective than "traditional studying" for knowledge retention.<sup>17</sup> Many medical schools have recently added frequent regular testing during the basic science years, and students and residents now commonly prepare for exams by using question banks. Feedback after questions further enhances learning and retention.<sup>17,18</sup> Finally, there is evidence that the enhancement effect of assessment works both with rote information and more complex skills<sup>19</sup> such as clinical reasoning.

These findings drive ABFM's goal to leverage assessment to support learning and keeping up to date. We aim to help Diplomates identify gaps in knowledge so that they can turn to the AAFP and other sources for focused education. FMCLA will be a mainstay of our efforts. As described in a recent editorial,<sup>20</sup> the FMCLA pilot has demonstrated significant learning among participants. Across over 11,000 Diplomates enrolled in the pilot, 95% used references to answer questions and 84% sought information after completing the test. Eighty-two percent indicated that they had changed their practice as the result of participating in FMCLA, and 89% reported incorporating FMCLA into their usual approach to keeping up with medical knowledge.

Interestingly, stakes seem to matter. Internal data allows us to contrast performance in over 11,000 Diplomates in the FMCLA pilot with the performance of over 20,000 Diplomates on Continuous Knowledge Self Assessment (CKSA). The 2 activities use a similar pool of questions, have the same IT platform, require 25 questions in a quarter and have the same

psychometric scale. Diplomates are more likely to have had a higher percent of correct answers on the higher-stakes FMCLA than the lower-stakes CKSA. This is also true for the 4,701 Diplomates who have done both activities simultaneously. Assessment used for summative evaluations of knowledge thus seems to serve a valuable role in facilitating learning.

## Conclusion

In service to the public, ABFM will continue to provide independent assessment of the knowledge necessary for Board-certified family physicians of their careers. At the same time, we will support continuous learning and keeping up to date. Linking assessments to board certification supports learning. More broadly, working with the AAFP and other partners, our role is to guide self-learning: to provide objective information to Diplomates about gaps of knowledge to help them target their CME. In addition, as part of our commitment to advancing the scientific basis of Board certification, we are collaborating with the American Board of Internal Medicine and the American Board of Medical Specialties to commission an extensive independent review of the evidence undergirding Board certification.<sup>21</sup> This comprehensive review will address the foundations of what we do: Do cognitive skills need to be kept current? Is self-assessment of knowledge sufficient? Does testing enhance learning? Do consequences matter? Finally, it is important to keep in mind that knowledge is only one dimension of excellence in clinical care, along with professionalism, the personal commitment to keep up to date and to improve care. Board Certification in Family Medicine is, and always has been, about more than knowledge.

Warren P. Newton, MD, MPH, American Board of Family Medicine, Department of Family Medicine, University of North Carolina, Thomas R. O'Neill, PhD, and David W. Price, MD, American Board of Family Medicine

## References

- Kruger J, Dunning D. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *J Pers Soc Psychol*. 1999;77(6):1121-1134.
- Zell E, Krizan Z. Do people have insight into their abilities? A meta-synthesis. *Perspect Psychol Sci*. 2014;9(2):111-125.
- Davis DA, Mazmanian PE, Fordis M, Van Harrison R, Thorpe KE, Perrier L. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *JAMA*. 2006;296(9):1094-1102.
- Choudhry NK, Fletcher RH, Soumerai SB. Systematic review: the relationship between clinical experience and quality of health care. *Ann Intern Med*. 2005;142(4):260-273.
- Anastasi A. *Psychological Testing*, 5th ed. Macmillan Publishing Co Inc; 1982.
- Raymond MR. A practical guide to practice analysis for credentialing examinations. *Educ Meas*. 2002;21(3):25-37.
- O'Neill TR, Peabody MR, Stelter KL, Puffer JC, Brady JE. Validating the test plan specifications for the American Board of Family Medicine's Certification Examination. *J Am Board Fam Med*. 2019;32(6):876-882.
- Cizek GJ, Bunch MB. *Standard Setting: A Guide to Establishing and Evaluating Performance Standards on Tests*. Sage Publications; 2007.
- Holland PW, Thayer DT. Differential item performance and the Mantel-Haenszel procedure. In: Wainer H, Braun H, eds. *Test Validity*. Lawrence Erlbaum Associates, Inc; 1988:129-145.
- O'Neill TR, Peabody MR, Puffer JC. The ABFM begins to use differential item functioning. *Ann Fam Med*. 2013;11(6):578-579.
- AERA, APA, NCME. *Standards for Educational and Psychological Testing*. American Educational Research Association; 2014.
- National Organization for Competency Assurance. *Standards for the Accreditation of Certification Program*. National Organization for Competency Assurance; 2004.
- O'Neill TR, Wang T, Newton WP. The American Board of Family Medicine's 8 years of experience with differential item functioning. *J Am Board Fam Med*. Under review.
- Ebbinghaus H. *Memory: A Contribution to Experimental Psychology*. Teachers College, Columbia University; 1885.
- Murre JMJ, Dros J. Replication and analysis of Ebbinghaus' forgetting curve. *PLoS One*. 2015;10(7):e0120644.
- Schwartz BL, Benjamin AS, Bjork RA. The inferential and experiential bases of metamemory. *Curr Dir Psychol Sci*. 1997;6(5):132-137.
- Larsen DP, Butler AC, Roediger HL III. Test-enhanced learning in medical education. *Med Educ*. 2008;42(10):959-966.
- Butler AC, Roediger HL III. Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Mem Cognit*. 2008;36(3):604-616.
- Phillips JL, Heneka N, Bhattarai P, Fraser C, Shaw T. Effectiveness of the spaced education pedagogy for clinicians' continuing professional development: a systematic review. *Med Educ*. 2019;53(9):886-902.
- Newton W, Baxley E, O'Neill T, Fain R, Rode K, Stelter K. Family medicine certification longitudinal assessment becomes permanent. *J Am Board Fam Med*. In press.
- Fraundorf S. Examining the foundational science behind continuing board certification. Presented at Annual American Board of Medical Specialties Virtual Conference, September 23, 2020.