BOARD NEWS

Family Medicine Certification Longitudinal Assessment after One Year

Warren P. Newton, MD, MPH, Kevin Rode, BA, Tom O'Neill, PhD, Roger Fain, BA, Elizabeth Baxley, MD, and Lars Peterson, MD, PhD

On January 4, 2019, the American Board of Family Medicine (ABFM) launched the Family Medicine Certification Longitudinal Assessment (FMCLA) pilot as an alternative to the 1-day Family Medicine Certification examination. Although FMCLA is intended as a "summative" assessment (ie, evaluating whether a candidate has the cognitive expertise necessary to be a board-certified family physician), a major reason we developed this alternative was to provide an experience that was also formative—one that helps family physicians identify gaps in their own knowledge and supports learning.[1,2] In the pilot, physicians eligible for selecting FMCLA are those who are in their 10th year of their certification cycle and are due to take the examination. Over time, we expect that everyone will have the opportunity to "roll into" this option coinciding with their examination year.

Participants in the Longitudinal Assessment answer 25 questions online every quarter, at a time and a place best for the physician, with access to any resource materials. The overall design builds in flexibility for adjustment to life circumstances, such as pregnancy, major illnesses, or family illness. A total of 300 questions over the 4-year period is necessary for making the final pass/fail judgment. The process could be completed in 3 years if a diplomate went straight through 25 questions per quarter, or she/he could take from 1 to 4 quarters off after the

first year of participation. As long as there is meaningful participation, diplomates remain certified throughout the 4 years of participation. The Longitudinal Assessment examination will be scored like the 1-day examination, and participants begin to receive individual performance reports at the end of the first year to guide them as they continue their questions. We expect a passing rate comparable to the 1-day examination, but those who fail will have an opportunity to take the 1-day examination to remain certified.

After 1 year of the pilot, what do we know? Clearly, Longitudinal Assessment is popular: 71% of diplomates chose it over the over the 1-day examination. The 2 groups (Longitudinal Assessment vs 1-day examinees) were similar in terms of age, sex, and type, size, and scope of practice. Interestingly, both groups cited convenience as their reason for making their choice of assessment methods. Overall, nearly 25% of diplomates preferred the 1-day examination; if current patterns hold up, this would represent more than 23,000 physicians! Given the numbers, ABFM is committed to continuing to offer diplomates a choice of methods if feasible.

Over 98% of pilot enrollees (n = 6009) finished the first year and continued into the second year of participation. In surveys, nearly all (>95%) were satisfied with their experience—the performance of the technical architecture, ability to track their status, and approving FMCLA as an alternative to a 1-day examination. The most common initial concern was the size and location of the clock, which we adjusted at the end of the first quarter. In addition, about 5% of participants initially reported that they did not have enough time to answer questions. In tracking the elapsed time, however, the average time per question across all participants

From American Board of Family Medicine, Lexington, KY (WPN, KR, TO, RF, EB, LP); University of North Carolina at Chapel Hill (WPN).
Conflict of interest: The authors are employees of the ABFM.
Corresponding author: Warren P. Newton, MD, MPH, American Board of Family Medicine, 1648 McGrathiana Pkwy, Ste 550, Lexington, KY 40511-1247 (E-mail: wnewton@theabfm.org).

was 2 minutes, 21 seconds. We believe that the 5-minute time limit is a reasonable proxy for actual practice, given the constraints of a usual 15 to 20 minute office visit managing multiple conditions; also, 5 minutes represents more than 3 times the average amount of time per question on the 1-day examination. Conversations with some participants suggest that learning how to use evidence-based resources quickly is an important issue. If this conclusion is supported in our final analysis, we believe it has important implications for residency training and continuing medical education.

We are starting to analyze data on how participants are experiencing Longitudinal Assessment. The most prevalent comment was that they were learning as they were going—that participation was actually changing their practice! In that sense, we have achieved the goal of having summative assessment also be formative. A total of 61% percent reported that they looked up information more than half the time while answering questions. More importantly, 75% reported seeking information on a clinical topic that was part of Longitudinal Assessment after taking the test. Although further experience is necessary, we believe that there is good evidence that Longitudinal Assessment is supporting continuous learning. In the coming months, we will explore further whether this alternative has had any impact more broadly on how participants are keeping up to date. Our long-term goal is to empower diplomates to identify their knowledge gaps and then, if appropriate, seek targeted continuing medical education to fill those gaps from the American Academy of Family Physicians, state chapters, or our other educational partners. In addition, as a part of making the results of our examinations more useful to physicians, ABFM is developing and validating a new examination blueprint that we hope will support more effective learning.[3]

A key question is the issue of whether Longitudinal Assessment allows ABFM to set a standard for the cognitive expertise necessary to be a board-certified family physician—does it perform comparably to the 1-day examination?[4] By design, FMCLA measures the same content and breadth of family medicine as the 1-day examination. After 1 year, it seems clear that item hierarchy—the difficulty of individual items compared with each other—is maintained. Psychometrically, this is 1 of the requirements for being able to set a standard fairly across the 2 formats. More experience with the

second year of the pilot will add precision to our statistical estimates, and comparisons between Longitudinal Assessment and the 1-day examination in various years will help us analyze the performance of FMCLA with the 1-day examination. Finally, a separate study will address the stability of individual performance over time.

What have we learned about the process? The ABFM examination is difficult—only about 60% correct is usually needed to pass, and it may be helpful for people to understand this. The corollary is that participants in the FMCLA should not get discouraged when they get 2 or 3 questions in a row wrong in any given quarter. In addition, we have valued the real-time feedback from diplomates about individual questions, which has identified questions with unclear language, raised important issues of clinical context, or prompted updating based on very new evidence. We replaced a few questions in year 1 as a result of this "crowd sourced feedback." Another concern ABFM has had is the potential risk to security of questions, which is more challenging in an online environment. Security is essential for fairness. So far, however, there have been no significant problems. We continue to monitor this issue actively, starting with diplomates' professionalism, including the commitment not to share questions with others. Finally, despite our efforts to differentiate between the Continuous Knowledge Self-Assessment questions (a 25-question per quarter self-assessment requirement option) and Longitudinal Assessment, some confusion remains among those who are participating in both, as well as in those who are not yet eligible to select the Longitudinal Assessment alternative because it is not their examination year. We will continue to work to differentiate the 2 activities.

An important consequence of the popularity of Longitudinal Assessment has been to reduce our ability to pretest examination questions. Pretesting items allows us to identify those multiple-choice questions that perform poorly psychometrically before they are used for standard setting. ABFM pretests all its multiple-choice items; this is part of our commitment to psychometric rigor. To reduce the need for pretested questions, the ABFM has implemented a series of actions, including eliminating the choice of modules in the 1-day examination. This frees up some of the pretested questions for use on FMCLA. Previous research has established that, perhaps paradoxically, selection of a module

(or 2) does not improve the examination scores or the passing rate. In addition, given our ongoing interest in making the content of the examination as relevant to practice as possible, we are now exploring how to develop some degree of customization to practice within Longitudinal Assessment, while retaining a standard of having items that represent a broad scope of family medicine practice.

What are next steps? In the summer of 2019, the ABFM extended the pilot to 2020. As we write this, participants in the Longitudinal Assessment pilot are receiving personalized reports that provide diplomates significantly more feedback than on the previous 1-day examination. This includes a summary of the "testing points" for questions they got wrong, along with additional reports linking testing points to questions participants considered relevant to their practice and for which they were confident—and got wrong. Our hope is that more specific feedback will help engage participants and help them target learning opportunities in a more focused manner. Pilot participants also received a visual "speedometer" that depicts how likely they are to pass/fail the examination, if their future performance is similar to that of the initial year of questions. We hope that this will be reassuring for most—suggesting that they just need to continue what they are doing to keep up to date. But, for some, this feedback may also serve as an "early warning" signal for being at risk of failing, thus providing them the opportunity to work on key areas of knowledge and to improve their performance as the assessment moves forward. Subsequent updates to this "How am I doing?" report will be provided quarterly throughout their time in Longitudinal Assessment—an improvement over the 1-day examination, which does not provide any ongoing feedback. Finally, we await additional evidence about the summative validity of the examination. This will take another 12 to 24 months.

Thank you for your engagement in the pilot—and in helping to guide its development. By working together, we can continually enhance the value of Family Medicine Certification.

To see this article online, please go to: http://jabfm.org/content/33/2/344.full.

References

1. Newton WP. The American Board of Family Medicine: what's next? J Am Board Fam Med 2019;32:282–4.
2. Newton WP, Rode K, O'Neill T, Fain R, Baxley E. Longitudinal assessment: where we are and why it is important. J Am Board Fam Med 2019;32:448–50.
3. Fain R, Newton WP, O'Neill TR. Creating a new blueprint for ABFM examinations. Ann Fam Med 2019;17:562–4.
4. O'Neill TR, Newton WP, Brady JE, Spogen D. Using the family medicine certification longitudinal assessment to make summative decisions. J Am Board Fam Med 2019;32:951–3.