Development and Results of the First ABMS Subspecialty Certification Examination in Sleep Medicine

Stuart F. Quan, M.D.1; Richard B. Berry, M.D.; 2; Daniel Buysse, M.D.; 3; Nancy A. Collop, M.D.4; Madeleine Grigg-Damberger, M.D.5; Susan M. Harding, M.D.6; Conrad Iber, M.D.7; W. Vaughn McCall, M.D.8; Michael J. Sateia, M.D.9; Stephen H. Sheldon, D.O.10; Michael H. Silber, M.D.,Ch.B.11; Adam Sorscher, M.D.12; Sally L. Davidson Ward, M.D.13; Sigrid Veasey, M.D.14; B. Tucker Woodson, M.D.15; Brian Hess, Ph.D.15; Roberta Kangilaski15

1Division of Sleep Medicine, Harvard Medical School, Boston, MA; 2Arizona Respiratory Center, University of Arizona, Tucson, AZ; 3Division of Pulmonary, Critical Care, and Sleep Medicine, University of Florida, Gainesville, FL; 4Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA; 5Department of Medicine, Johns Hopkins University, Baltimore, MD; 6Department of Neurology, University of New Mexico, Albuquerque, NM; 7UAB Sleep-Wake Disorders Center; University of Alabama at Birmingham, Birmingham, AL; 8Pulmonary and Critical Care Medicine, Hennepin County Medical Center; University of Minnesota, Minneapolis, MN; 9Department of Psychiatry and Behavioral Medicine, Wake Forest University, Winston-Salem, NC; 10Section of Sleep Medicine, Dartmouth Medical School, Lebanon, NH; 11Sleep Disorders Center, Children’s Memorial Hospital, Chicago, IL; 12Department of Neurology and Center for Sleep Medicine, College of Medicine, Mayo Clinic, Rochester, MN; 13Department of Community and Family Medicine, Dartmouth Medical School, Hanover, NH; 14Division of Pediatric Pulmonology, Children’s Hospital Los Angeles, and Keck School of Medicine, University of Southern California, Los Angeles, CA; 15Center for Sleep and Neurobiology and Department of Medicine, University of Pennsylvania School of Medicine, Philadelphia, PA; 16Department of Otolaryngology and Human Communication, Medical College of Wisconsin, Milwaukee, WI; 17American Board of Internal Medicine, Philadelphia, PA

In November 2007, the first Certification Examination in Sleep Medicine was administered to 1,882 candidates under the cosponsorship of five member boards of the American Board of Medical Specialties (ABMS)—the American Board of Internal Medicine, the American Board of Family Medicine, the American Board of Otolaryngology, the American Board of Pediatrics, and the American Board of Psychiatry and Neurology. The pass rate was 73%. This paper chronicles the history of a certification examination in Sleep Medicine and the development of this new ABMS examination.

Citation: Quan SF; Berry RB; Buysse D; Collop NA; Grigg-Damberger M; Harding SM; Iber C; McCall WV; Sateia MJ; Sheldon SH; Silber MH; Sorscher A; Ward SDL; Veasey S; Woodson BT; Hess B; Kangilaski R. Development and results of the first ABMS subspecialty certification examination in sleep medicine. J Clin Sleep Med 2008;4(5):505-508.

The Certification Examination in Sleep Medicine was administered in November 2007 for the first time under cosponsorship of five member boards of the American Board of Medical Specialties (ABMS)—the American Board of Internal Medicine, which is the designated administrative board, the American Board of Family Medicine (ABFM), the American Board of Otolaryngology (ABOto), the American Board of Pediatrics (ABP), and the American Board of Psychiatry and Neurology (ABPN). Preceding this administration was a 28-year history of testing by the American Board of Sleep Medicine (ABSM), beginning in 1978 with an examination in clinical polysomnography, passed by 21 candidates, and ending with the 2006 ABSM examination. During the 28-year period ending with the 2006 ABSM examination, 3,445 individuals were certified as diplomates of the ABSM.

For the 1978 examination, candidates answered written true-false questions and oral questions posed by an examiner. In 1980, candidates completed a two-part examination given on separate dates. The first part consisted of 3 sections of multiple-choice questions. Candidates who successfully passed the first part took the second part, in which they completed essays based on interpretation of polysomnograms and multiple sleep latency tests. By 1991, the second part had evolved into an examination that required short written answers instead of essays, and a computer-based presentation of polysomnogram samples in lieu of paper recordings was introduced in 2002. Eventually, in 2005, a single-day computer-administered examination consisting only of multiple-choice questions was utilized. This format was very similar to that used for the 2007 Certification Examination in Sleep Medicine.1

The American Academy of Sleep Medicine and the ABSM prepared the way for certification by ABMS member boards, applying for recognition of sleep medicine training by the Accreditation Council for Graduate Medical Education (ACGME) and taking a major role in seeking ABMS approval of the certification process.1 Accreditation requirements for sleep medicine training programs were established in 2004, and the ABMS approved sleep medicine certification in 2005, leading to development of the 2007 Certification Examination in Sleep Medicine.
When the Sleep Medicine Test and Policy Committee (see acknowledgments) convened for the first time in 2005, one of its major tasks was to establish the content outline, or blueprint, for the examination using the ABSM content outline as a starting point (Table 1 presents the primary medical content areas of the examination). Next, the committee developed topic assignments in each content area and received in-depth education from ABIM experts in the preparation of well-structured test questions. Individual committee members wrote questions in multiple content areas; this cross-disciplinary approach was intended to ensure appropriate balance among the disciplines represented by the cosponsoring boards. Question content focused on clinical abilities needed by physicians practicing sleep medicine, regardless of their primary specialty. Questions were rigorously reviewed and critiqued in multiple meetings of the entire committee. Once questions were discussed, revised, and rewritten, the committee made its final selection of questions for the examination. The ABIM website contains further information about the test blueprint (http://www.abim.org/pdf/blueprint/sleep_cert.pdf) as well as the test development process (http://www.abim.org/about/examInfo/developed.aspx).

The Certification Examination in Sleep Medicine contained 240 single-best-answer multiple-choice questions and was computer administered at more than 200 test centers across the United States and other countries. After the examination was administered, the statistical performance of all questions was reviewed to ensure accuracy of both the content and the answer for each question. Comments about specific questions provided by candidates in a survey administered at the end of the examination were also considered. This process is called key validation. For questions judged to be unfair (e.g., affected by recent changes in sleep medicine) or flawed in any way, the committee had the option of scoring multiple answers correct during the final scoring process. Scoring results were changed for approximately 4% of the questions during the key validation process, rather than the more typical range of 15% to 20% for new examinations.

All candidates were held to the same absolute content-based standard for passing the examination, rather than a relative standard that is dependent on their performance in comparison with other candidates. It was thus theoretically possible that all candidates, or conversely no candidates, could pass the examination. An absolute minimum passing score was established by the committee using the modified Angoff method, which is a multistep judgment-based process. All 15 committee members representing the 5 boards cosponsoring the examination participated in the standard-setting process. The committee first discussed the knowledge and skills that minimally qualified, or “borderline,” examinees should possess. Next, all questions were reviewed one at a time and each committee member provided an estimate of the percentage of borderline candidates who would correctly answer the question. Estimates for each question were displayed for the group to review, and members whose estimates differed significantly from the group’s average were asked to offer a rationale to support their estimate. Discussion followed, with all members given the opportunity to change their estimates. Final judgments were then recorded, and the average percentage for each question was calculated. These average percentages were translated to proportions, and the sum of these proportions comprised the minimum number of questions that candidates had to answer correctly to pass the examination.

To help evaluate the results of the modified Angoff process, a separate independent standard-setting procedure was conducted using the Hofstee method, which combines features of both relative and absolute standard setting. That is, judges are required to define minimum and maximum acceptable passing scores and fail rates, which are then analyzed with examination performance data. Specifically, committee members’ minimum/maximum acceptable passing scores and fail rates were averaged, and these four mean values were plotted on the cumulative frequency distribution of examination scores. A viable standard was identified as the score where the distribution intersects with the midpoint of the averaged minimum/maximum passing score and fail rate.

Once the standard-setting process was complete, the committee reviewed and approved the final results. The minimum passing score for the examination was 170 correctly answered questions (70.8%). The minimum passing score determined from the Hofstee process was identical to that derived from the Angoff method. Raw scores were standardized for reporting purposes, using a scale with a mean of 500 and a standard deviation of 100. The minimum passing score on the standardized score scale was 442.

## Examination Performance and Results

A total of 1,882 candidates took the examination. There were 1,228 (65%) ABIM candidates, 460 (25%) ABPN candidates, 83 (4%) ABP candidates, 78 (4%) ABOto candidates, and 33 (2%) ABFM candidates. The overall percentage of candidates passing the examination was 73%.

Three admission pathways to the examination were available to prospective candidates. Pathway A was self-attestation, subject to possible audit, of the equivalent of 12 months of full-time, post-training practice experience providing clinical care to patients with sleep disorders, accumulated over a maximum of 5 years prior to examination application. Candi-
There was nearly universal agreement that the testing environment. Overall, 88% of candidates completed the survey. Security, presentation of the examination content, and the testing industry standards. The reliability and reproducibility values meet the same pass/fail decision if repeatedly tested with equivalent examinations. The coefficient α for the examination was 0.91. This value indicates that the variability in scores is largely due to differences in the true abilities of the candidates. The reproducibility of the pass/fail decision, which is related to score reliability, is an estimate of the proportion of candidates who would receive the same pass/fail decision if repeatedly tested with equivalent examinations. The coefficient α for the examination was 0.91. This value indicates that the variability in scores is largely due to differences in the true abilities of the candidates. The pass/fail reproducibility for the examination was 0.89, which indicates that approximately 89% of the candidates would receive the same pass/fail decision if retested with an equivalent examination. The reliability and reproducibility values meet testing industry standards.4

Finally, data were analyzed from a computer-based survey administered to examinees at the testing centers immediately after they completed their examinations. The survey asked about their satisfaction with various aspects of their testing experience, including satisfaction with registration procedures, test center security, presentation of the examination content, and the testing environment. Overall, 88% of candidates completed the survey. There was nearly universal agreement that the testing environment was comfortable (92%) and that the testing system was easy to use (99%). Most agreed that the question text was easy to read (81%) and that the illustrations and figures were clear (75%). The vast majority of the candidates were satisfied with their overall testing experience (91%). These data indicate that the delivery of the examination via computer was of high quality.

**SUMMARY**

The data presented here indicate that the first ABMS Certification Examination in Sleep Medicine performed very well. Examination scores were reliable and the resulting pass/fail decisions were reproducible. Furthermore, the overall pass rate was judged to be reasonable, and the performance of each of the three admission pathways was as expected.

The Sleep Medicine Test and Policy committee will continue to expand and improve the pool of questions for the next test, which will be given in October 2009. Through the collegial effort of the committee, the examination will continue to serve as a high-quality assessment instrument to identify physicians who possess the medical knowledge required to be certified in sleep medicine.

**ACKNOWLEDGMENTS**

Representatives of each of the cosponsoring boards on the Sleep Medicine Test and Policy Committee are the following: ABIM (Drs. Quan [Committee Chair], Berry, Collop, Hard- ing, Iber, Vasey [Physician Secretary]); ABFM (Dr. Sorscher); ABOto (Dr. Woodson); ABP (Drs. Sheldon and Davidson Ward); ABPN (Drs. Buysse, Grigg-Damberger, McCall, Sateia, Silber). Dr. David White (ABIM) was the chair and member of the committee during its first 2 meetings in 2005.

**DISCLOSURE STATEMENT**

This was not an industry supported study. Dr. Berry has received research support from Itamar and Respiromedics. Dr. Buysse serves as a consultant for Actelion, Arena, Cephalon, Eli Lilly, GlaxoSmithKline, Merck, Neurocrine, Neurogen, Pfizer, Respiromedics, Sanofi-Aventis, Sepracor, Servier, Somnus Therapeutics, Stress Eraser, Takeda, and Trancept Pharmaceuticals. Dr. McCall has received research support from Sepracor, Sanofi-Aventis, and MiniMitter and had consulted for Sepracor, Somaxon, and Sealy. Dr. Hess and Roberta Kangilaski are both full-time employees of the American Board of Internal Medicine, the administrative board for the examination that is the subject of this paper. The other authors have indicated no financial conflicts of interest.
REFERENCES


