Audit Panel Report of the July 2003 & 2004 NBPME Part I Examination

Prepared by:

Patrick Jones, PhD, Audit Panel Chair
Mark Raymond, PhD
Gerald Rosen, EdD

Audit Panel Report of the July 2003 & 2004 NBPME Part I Examination

Introduction

In July 2006, the Council of Deans of the American Association of Colleges of Podiatric Medicine (AACPM) and the Board of Directors of the National Board of Podiatric Medical Examiners (NBPME) convened a three-member panel to conduct an independent audit of the July 2003 and July 2004 NBPME Part I Examination. The panel included Dr. Gerald Rosen, a measurement psychologist appointed by the AACPM, Dr. Mark Raymond, a measurement psychologist appointed by the NBPME, and Dr. Patrick Jones, a measurement psychologist selected by Drs. Rosen and Raymond. Each panel member has extensive training and experience in psychometrics and with the design, development, administration, scoring and evaluation of credentialing examinations. Dr. Jones served as the Chair of the panel.

The panel used an audit protocol provided by the AACPM and NBPME to examine the validity and reliability of scores from the July 2003 and July 2004 administrations of the NBPME Part I Examination. Beginning with the July 2003 testing cycle, the methods used to develop and administer the exam shifted from a sequential mastery test administered by computer to a fixed-form, linear test administered in a paper-and-pencil format. Coincident with these changes, the passing rate for first-time candidates decreased. The audit sponsors requested that the panel investigate the factors that caused this decrease in passing rate.

As part of the audit, the panel requested access to several sources of information about the Part I Examination. These included primary documents relating to the development, administration, scoring and analysis of the exam during the period of interest. In addition, the panel received presentations from NBPME and Thomson Prometric representatives on issues pertaining to the psychometric characteristics of the test. The NBPME contracts with Thomson Prometric for exam services in support of its testing programs. At the conclusion of these presentations, the panel had an opportunity to question NBPME and Thomson Prometric representatives on matters germane to the focus of the audit. The panel was very satisfied with the depth of the information provided by the NBPME and Thomson Prometric and their responses to panel questions.

This report is organized according to the five (5) general issues of concern identified by the AACPM and the NBPME. For each issue, the panel offers conclusions based on its review of available information and recommendations to improve the quality of the testing program. Responses to each of the questions posed in the audit protocol are provided in Appendix 1. Readers of this report are directed to the *Bulletin of Information* published by the NBPME for a description of the content and format of the Part I Examination.

Issue #1: Were the scores obtained from the July 2003 and July 2004 administrations of the Part I Examination valid and reliable according to industry standards?

See comments below *Issue #2*.

Issue #2: Is the examination process, from job analysis to the NBPME Board's determination of cut scores, consistent with industry standards?

The procedures used to develop, administer and score the Part I Examination were consistent with industry standards given the design of and goals for the testing program. The test specifications appeared to be based, for the most part, on the results of a well-designed job analysis study, and care was taken to ensure that the July 2003 and July 2004 forms of the Part I Examination were constructed to meet the test blueprint based on the outcomes from the job analysis study. The methods used to develop and review test items were sound, and input was gathered from subject matter experts who were familiar with the content areas of the test blueprint and the educational preparation of the candidate population along with item development specialists from Thomson Prometric. Construction of the test forms was based on the test specifications and care was taken by Thomson Prometric to ensure that item cueing was avoided. The test was administered in a standardized and secure fashion, and test results were scored in a manner designed to promote accuracy and rigor. The reporting of test scores was consistent with industry practices. The procedures used to generate criterion-referenced passing scores were properly designed and conducted, and the NBPME Board was provided with sufficient information to establish cut scores for each test form.

An evaluation of the item and test statistics indicated that the reliability of test scores was sufficient for the purpose of the examination program. The validity of test scores was generally advanced by the methods used to develop, administer and score the exam.

One important aspect of score validity pertains to the appropriateness of decisions based upon test scores. For the Part I Examination, the NBPME relied on the results from criterion-referenced, passing score studies conducted for each administration of the test to establish and maintain reasonable performance standards over time. Concerns regarding this approach are presented under *Issue #4*.

Panel Recommendations:

- The job analysis report indicated that numerous responsibilities offered by the Job Analysis Task Force which failed to exceed the importance rating-scale threshold established for excluding elements of the job analysis from further consideration (i.e., an average value of 2.5 based on survey sample results) were still included on the final list of responsibilities. Although the Task Force should have some latitude in making these types of decisions, the survey sample's input should receive the greatest weight in determining test content and the survey results should be overruled in only certain, limited circumstances.

- The development of more complex item types should be considered. These item types might include multiple, correct answer formats with more than four options. The use of complex item types would promote test security and enhance the psychometric quality of the test.

- In addition to contracting with faculty from podiatric medical colleges to write items, test items should also be solicited from practitioners and licensing board members.

- Providing item writers with item shells and using techniques to "clone" items in the basic sciences would lead to larger item yields and more efficient item development.

- Item review and approval should be informed by requiring subject matter experts to provide ratings of the criticality and importance of item content to safe practice as an entry-level podiatrist.

- Consideration should be given to lengthening the Part I examination by fifteen (15) items. After this one hundred and sixty-five (165) item test has been administered and scored, fifteen (15) items would be removed from the test based on a review of item analysis data and challenges submitted by candidates. In addition, the final set of one hundred and fifty (150) items should closely approximate the content requirements of the test specifications. A candidate's passing status would be based on his/her performance on the final set of test items after the fifteen (15) flawed items have been removed. Adoption of this procedure would improve the psychometric quality of the exam by eliminating the need to grant candidates credit for flawed test items and by ensuring that the content of the final set of test items closely matches the test specifications.

- Subject matter experts who serve as judges during standard setting studies should complete a questionnaire that documents their understanding of and confidence in the procedures used to set cut scores for the exam.

Issue #3: Would the examination's validity be enhanced by permitting faculty from each of the colleges of podiatric medicine to review the entire item bank?

Faculty members from the colleges of podiatric medicine currently play a significant role in the development of examination items. The panel does not recommend convening faculty from each of the colleges of podiatric medicine to review the entire item bank, because this type of review is neither an efficient nor cost-effective means of improving the validity of an examination. However, the panel does recommend two changes to the current item and exam review protocol.

Panel Recommendations:

- Faculty from colleges of podiatric medicine who are subject matter experts in the content areas of the Part I examination should be added to the review panels currently convened to evaluate new items. To minimize item exposure, it is proposed that faculty from any particular podiatric medical college be limited to a single review panel during an annual item review cycle. If this recommendation were adopted, a rotation schedule could be implemented to ensure that each college would eventually contribute faculty to the item review panels for each content area of the test.

- The quality of the test would be enhanced by having a broader group of subject matter experts appointed by the NBPME (e.g., practitioners, licensing board members, podiatric medical college faculty) collaborate with Thomson Prometric staff to oversee the assembly of final forms of the exam. To determine the composition of this test review committee, the panel suggests that the NBPME first identify stakeholder groups who have a legitimate interest in the outcomes from the Part I examination. Once these stakeholder groups have been determined, a committee selection matrix could be developed that includes as a separate dimension other important factors that should inform the final composition of the test review committee (e.g., area of expertise, years of experience, gender, race).

Issue #4: Could a significant proportion of the variability in the pass rates on the July 2003 and July 2004 Part I Examination been due to factors intrinsic to the tests?

Variation in pass rates observed for the administration of the July 2003 and July 2004 Part I Examination may have been due to the following factors. These factors include the composition and psychometric characteristics (e.g., difficulty and discrimination) of the test forms, the administration of only new items on the test that eliminated candidates' prior access to test content, the cut scores used to determine passing status, and changes in the ability level of the candidates. Candidate ability is influenced by their suitability for an educational program in podiatric medicine, their level of preparation for the Part I Examination, and other candidate-specific factors (e.g., motivation, persistence, confidence).

Based on the information made available to the audit panel, it was not possible to answer this question definitively. Although the panel concluded that the procedures used to develop, administer and score the Part I Examination were designed and implemented to minimize the impact of factors intrinsic to the tests on the variability of pass rates, it is quite possible that some of the variation in pass rates can be attributed to unintended shifts in the standard of performance required to pass the exam. That is, it is quite possible that the pass/fail outcomes for certain candidates in 2003 and 2004 were determined in part by the particular form of the test administered to them.

It is not surprising to find significant variations in cut scores and passing rates for the July 2003 and July 2004 testing periods. The Part I Examination was comprised of all new items and the cut scores were based on the pooled judgments of subject-matter experts (i.e., Angoff ratings). The Angoff standard-setting procedure – in the absence of other test development, score equating, and score scaling techniques – is not adequate to assure the comparability of pass/fail standards from one year to the next, and it is not an effective procedure to construct a score scale that reflects a consistent level of ability over time. Having noted these issues, the panel also acknowledges that security breaches observed in 2002 most certainly interfered with the ability of the NBPME to maintain comparable pass/fail standards and a consistent score scale through the use of equating or linking methods.

Panel Recommendations:

- Test development and standard setting activities should be transitioned to procedures that are common to models that feature equating or linking procedures to promote the comparability of test forms and cut scores. Item pretesting should be a component of these models. Based on representations made to the panel, it appears that the NBPME and Thomson Prometric plan to implement these test development methods in the near future.

- Until such time when equating or linking plans have been implemented, Thomson Prometric should consider additional procedures during the conduct of standard setting studies to improve the comparability of judges' ratings among different forms of the Part I Examination.

Issue #5: <u>If a significant proportion of the variation in pass rates during the administration of the July 2003 and July 2004 Part I Examination could be attributed to factors intrinsic to the tests, the audit panel shall make specific recommendations regarding measures that can be reasonably expected to minimize such intrinsic variability in the future.</u>

The reader is directed to the panel's recommendations presented earlier in this report for measures that should be considered to improve the quality and consistency of the Part I Examination.

General Recommendations to Improve Communication with Constituents and Foster an Atmosphere of Trust and Cooperation

During the course of the review, the panel noted that the NBPME does not share some types of information and reports with the public and its constituents that many certification and licensure boards routinely make available to interested parties. In the spirit of fostering an atmosphere of trust, the auditors suggest that the NBPME publish an annual report documenting the psychometric properties of its examinations, examinee volumes, the performance of first-time and repeat examinees, and, possibly, selected demographic characteristics of examinees. It is further suggested that such reports document examination and examinee performance over time.

It is also suggested that the NBPME and the educational programs in podiatric medicine engage in cooperative efforts to collect and share the types of student educational data that would contribute to the Board's efforts to conduct empirical investigations regarding the validity of its examinations. The availability of measures such as MCAT scores, performance ratings obtained during clinical rotations, course grades, and student demographics would provide NBPME and the schools with information that might be useful not only for conducting validity studies, but also for conducting studies that might inform educational practices and benefit the profession. As an example, a hypothesis that might reasonably be proposed to explain an increase or decrease in pass rates over time is that students are more or less qualified now than in the past. The availability of MCAT scores could help confirm or disconfirm such a hypothesis.[1]

It was clear from the documentation provided to the auditors that NBPME testing practices are generally quite sound. A final recommendation is to prepare and disseminate informational brochures that describe in detail the procedures used to develop, administer, and score examinations (including standard setting and answer key verification). In the auditors' experience, such documentation can help build public trust and support for the Board's activities. Two of the auditors (Drs. Jones and Raymond) would be pleased to share examples of these brochures.

---

[1] Although the suggestion in this footnote may not be viewed as being in the spirit of cooperation, it is consistent with fostering an atmosphere of academic honesty and integrity. The use of MCAT scores, in conjunction with the deliberate use of both new and used items, provides a useful way to investigate the extent to which used items are circulated among candidates and/or educational programs. If students from certain educational programs score higher on previously administered items than predicted by their performance on: (a) the new items on a test form and (b) their MCAT scores, then there would be reason to believe that students had prior access to test content. At least one medical specialty board routinely conducts such analyses as a means to encourage honest test-taking practices.

Appendix 1

Psychometrician Panel Responses to AACPM and NBPME Audit Protocol

AUDIT OF TEST DEVELOPMENT

1. **Is the test or evaluation instrument based on a job analysis?**

The NBPME Part I exam is based on the results from a job analysis study.

- Does the report include the survey, the results, recommendations, committee members and their qualifications?

Yes.

- If no job analysis, what is the basis for the test?

Not applicable.

2. **Do test specifications exist?**

The Part I exam specification are published in the *Bulletin of Information 2006 Part I and II Examinations* published by the NBPME.

- Are test specifications linked to the job analysis? By what process?

The NBPME Test Specifications Committee determined the composition of the Part I exam blueprint based on the results of a survey of podiatric medical doctors. As part of the survey, respondents were directed to provide weights for each major content area identified in the job analysis. The Committee referred to the weights provided by the survey sample and the allocation of test content on the then current exam to devise recommendations for the final test blueprint.

The survey sample also evaluated the knowledge areas and responsibilities assessed by the Part I exam, and the survey sample provided ratings of their importance and relevance to competent, entry-level practice as a podiatric medical doctor. The Committee was guided by the survey sample results during the final preparation of the test specifications.

- If no job analysis, how were the test specifications developed?

Not applicable.

- How was the test format (written, oral, practical) determined?

The purpose of the Part I exam is to assess whether an individual possesses the knowledge required to practice as a minimally competent entry-level podiatrist. Given this goal, the NBPME determined that a written test consisting of multiple-choice items was the most appropriate test format.

- How were dimensions of complexity or cognitive level established and supported?

The dimensions of complexity and cognitive level are not part of the Part I examination test specifications. Complexity and cognitive level are reflected in the item bank based on the development of items that address specific components of the areas of responsibility and knowledge delineated in the test specifications.

3. **Are the test items consistent with test specifications and cognitive level?**

Yes.

4. **How are the items developed?**

- Who are the item writers and what are their qualifications?

Item writers are drawn from faculty at colleges of podiatric medicine in the United States. These faculty members are subject matter experts in the content areas assessed by the Part I exam.

- How are the item writers selected?

Administrators at each college of podiatric medicine are asked to nominate individual faculty members based on their subject matter expertise.

- How are the item writers trained?

Thomson Prometric, the testing service under contract with the NBPME to oversee the development, administration, and scoring of the Part I exam, provides item writers with support during the item development process. All item writers are provided with an Item Development Manual prepared by Thomson Prometric to guide their item authoring activities.

- How are the items initially reviewed and approved?

Items are initially reviewed and approved by faculty from medical colleges in the United States who serve on item review panels under the direction of Thomson Prometric test development specialists.

- Is the review process objective and based on consensus of experts?

Yes.

- Is each item supported by a text reference or written documentation of some sort?

Yes.

- Is each item then linked with the test blueprint and job analysis?

Yes.

- Who has the authority to edit items?

Members of the item review panels and Thomson Prometric test development specialists have the authority to edit items.

5. **How are items introduced into production?**

- Are items pre-tested in some manner?

Because of test security concerns, items are not pre-tested prior to inclusion on an operational test form.

- Are item statistics evaluated against standards for item difficulty and discrimination?

Yes.

- Are results of item bias and adverse impact studies available?

No. Thomson Prometric is currently collecting data to permit the conduct of item bias and adverse impact studies. When a sufficient number of candidates exist, such studies will be completed.

6. **How are test forms assembled?**

- What is the percentage of old (used) items on the test?

No previously administered items are included on new test forms.

- What is the percentage of new items?

100%.

- Are pretest items included?

No.

- Who reviews final test form for content and correctness of answer key?

Subject matters experts appointed by the NBPME and test development experts from Thomson Prometric review the final test form for content coverage and answer key accuracy.

- How many forms of the test need to be developed for an administration or a specified time period?

A single unique form of the test is required for administration on each test date.

7. **Is the bank of items large enough?**

- Are there sufficient items to develop several examination forms?

Yes.

- How often are items re-used?

Never.

- Are there enough items in the bank to replace a stolen or compromised form?

Yes.

8. **How is the passing score established?**

- What method is used to establish a passing score?

A variation of the standard setting method often attributed to Dr. William H. Angoff from the Educational Testing Service is used to establish a passing score. A standard setting study is conducted for each new form of the Part I exam.

- How is the standard of competence defined?

The standard of competence is defined by subject matter experts working under the guidance of test development specialists from Thomson Prometric.

- Is it criterion-referenced or norm-referenced?

The standard of competence is criterion-referenced.

- Does the method incorporate the systematic judgment of a representative group of content experts?

Yes.

- How are the experts selected, and what are their qualifications?

The NBPME has devised a matrix of qualifications that guide the selection of subject matter experts who serve as judges during the standard setting process. Judges must be practicing members of the podiatric medical profession and the panel of judges is representative of the profession with respect to practice setting, experience, gender, race and other important demographic characteristics.

- What is the reasoning for the standard that was selected?

The NBPME Board of Directors selects the final passing score based on the results from each standard setting study and recommendations from the panel of judges and Thomson Prometric staff. The standard is intended to reflect the level of performance on the Part I exam equivalent to a minimally competent, entry-level podiatrist.

- Is the passing standard the same for all forms of the test?

No.

9. **How often is test content updated? What procedures are in place to assure currency of test content?**

Test content is updated annually. A new form of the test is administered on each test date. Item and test reviews are conducted each year, and subject matter experts are directed to attend to issues of content currency during these reviews.

10. **What procedures assure that test content is secure during the test development process? Consider office security, electronic security, and security measures for at-home writing of questions, if used.**

Thomson Prometric staff are responsible for ensuring that item and test content remain secure at all times during the development and review cycles. Items and exams under the control of Thomson Prometric are protected through a series of physical and electronic security and audit measures. Access is severely restricted to

only those staff and volunteer subject matter experts who have a legitimate reason to work with exam material. Item developers and reviewers are monitored to ensure that they comply with security and confidentiality protocols.

AUDIT OF STATISTICAL ANALYSES AND SCORING

1. **Are item statistics calculated?**

- Are the item statistics analyzed? What criteria are applied to keep or reject items?

Traditional Classical Test Theory indexes of item difficulty and discrimination are calculated after each administration of the Part I exam. Thomson Prometric staff applies consistent standards based on these indexes along with a consideration of comments from test candidates to identify items that may be flawed.

- What happens to items that are deemed inadequate or flawed?

Candidates receive credit for an item that is deemed inadequate or flawed.

- Are the statistics stored for use in item revision?

Yes.

2. **Are reliability statistics available?**

- On what sample group were the statistics calculated?

Reliability coefficients are based on the test performance of first-time candidates.

- Is the exam reliable enough to support making pass/fail decisions?

Yes.

3. **Is the test equated so that reported scores on different forms will be a result of candidate ability and not item difficulty?**

No. The NBMPE and Thomson Prometric conduct passing point studies and scaling procedures for each new test form. These procedures are designed to ensure that a test candidate's score and pass/fail status are not influenced by differences in the difficulty or discrimination of test forms.

- What method is used to equate the tests?

Not applicable.

- Why was that method selected?

Not applicable.

- On what sample group was the equating performed?

Not applicable.

4. **What are the pass/fail rates, and the sample groups on which they were calculated? Are the pass/fail rates reviewed for expected outcomes, given the sample group?**

Passing rates are calculated based on the performance of 1st time candidates, candidates who are repeating the test, and the total group of candidates tested on a particular test date. The NBPME and Thomson Prometric review the passing rates of each group for expected outcomes. The NBPME informed the panel of psychometricians that specific passing rate summaries could be made available to the AACPM upon request to the NBPME.

5. **How are tests scored? What quality control procedures assure scoring accuracy?**

Thomson Prometric uses scanning equipment designed to read pencil marks provided by test candidates on answer sheets to create an electronic file of item responses for each candidate. Quality control procedures consistent with industry standards are employed to make sure that the pencil marks provided by test candidates match the item responses contained in the electronic files. These electronic files of item responses are then compared against the examination answer key, and item responses are recoded to integer values representing correct and incorrect answers. Computer programs are then used to derive scaled and total scores, and pass/fail status is determined based on the application of the passing score approved by the NBPME.

During each phase of the scoring procedure, Thomson Prometric staff use quality control procedures and audit checks to make sure that scoring is completed with an extremely high level of accuracy.

6. **How are individual candidate scores reported?**

- Are they reported promptly?

Ordinarily, candidates receive test results six weeks after the test administration.

- Do they indicate what scoring rules are used?

Yes.

- Are candidate scores reported as subscale and total scores?

No. A scaled score is computed for a test candidate's performance on the total test. This scale ranges from approximately 55 to beyond 75, and the scaled score passing point has been set by the NBPME at 75. Candidates who pass the exam receive a score report with a "pass" indicated; failing candidates receive a report with a failing scale score between 55 and 74. Failing candidates also receive diagnostic indicators for the seven, basic science subjects included in the Part I exam. These diagnostic indicators state if the test candidate's performance was below the level of minimum competence or at or above the level of minimum competence.

- Are numeric scores issued or just pass/fail status?

See previous answer.

- If candidates fail, do they receive any diagnostic information?

See answer above previous answer.

7. **What types of reports are generated: school results, committee reports, etc.?**

Colleges of podiatric medicine receive summary statistics of the exam performance of their students relative to all 1st time candidates tested on a particular test date. Colleges of podiatric medicine may also receive the individual score report from a student attending a particular school if that student authorizes Thomson Prometric to release his or her score report. Test evaluation reports prepared by Thomson Prometric are sent to the NBPME at the conclusion of each test administration cycle.

8. **Are examination reviews or challenges allowed? If candidates can contest individual items on the test, what is the procedure for the protest? If not, what is the reason for not allowing appeals?**

Candidates with complaints or comments about test center facilities and/or supervision, examination content, or any other matter related to the testing program may complete a comment form at the test center or write to Thomson Prometric within one week of the test date. Thomson Prometric staff and the NBPME consider these complaints and comments during the scoring and reporting of exam results.

9. **What complaints have been expressed about the current test? What responses have been generated by the testing agency?**

Candidate complaints generally fall into two broad categories: complaints about the test administration conditions and challenges relating to item content or the accuracy of test scores. Thomson Prometric investigates complaints about testing conditions and takes appropriate actions in collaboration with the NBPME to address problems uncovered during these investigations (e.g., change test site or testing staff, retest all

candidates who failed the test at a particular test site). Challenges to item content and the accuracy of test scores are investigated by Thomson Prometric and the NBPME (see response to previous question).

ADDITIONAL QUESTIONS

1. **What resources are available to the test provider organization(s)?**

- What physical resources are available to the staff and consultants - computers, security, etc?

Thomson Prometric has adequate resources to manage the development, delivery, and scoring of the Part I examination. As noted in a previous answer, the security of the examination and related materials is maintained through a combination of procedures that are consistent with industry practices. Subject matter experts who serve as consultants to the NBME during the development of items and exams are selected based on their ability to perform the duties assigned to them and their willingness to adhere to strict security procedures.

2. **What is the financial status of the organization?**

- How is the testing program financed?

The Part I exam is funded on the basis of candidate fees for exam services.

- Is there a reasonable expectation of stability? (Revenues meeting expenses?)

Yes.

- Do examination fees fund other programs?

No.

- Are fees reasonable compared to other similar examinations? Are fees appropriate to the profession? Is justification provided when fees are increased?

The candidate fees for exam services are comparable to other similar licensure examination programs with testing volumes similar to the Part I exam. When fees are increased, the NBPME releases information to the public explaining the reasons behind the increase.

3. Is the bank of items large enough?

Yes. See answer to question # 7 under **AUDIT OF TEST DEVELOPMENT** section of this report.

- If the test is or was computer adaptive, are or were there enough items at each difficulty level to administer a test that can make an accurate assessment of an individual's level of proficiency?

Not applicable as the Part I exam is not administered as a computer adaptive test.

4. **If the test is now administered in one format, but is intended to be comparable to tests previously administered in other formats, is there data to demonstrate that the results are comparable?**

Yes.

5. What procedures are established to maintain security of the test materials?

See answer to question #10 under **AUDIT OF TEST DEVELOPMENT** section of this report.

- When the test was last computer-delivered, what security measures were in place and were established to monitor the testing site(s) and the candidates?

Thomson Prometric and NBPME used methods consistent with industry standards to maintain the security of the Part I exam and the validity of test scores.

- If there will be a return to computer-delivered tests, what security measures will be in place to monitor the testing site(s) and the candidates?

The Part I exam will be administered by computer at Thomson Prometric testing sites beginning in July 2007. Test center staff will verify the identity of and monitor candidates through direct visual observation and electronic measures. Candidates will not be allowed to bring unauthorized materials into the testing room, nor will they be able to remove any notes or other materials from the testing room. Thomson Prometric and the NBPME will monitor the security environment at test centers on a regular basis.

- How are examination materials returned to test developer if administration staff is not part of test development staff?

Thomson Prometric develops and administers the Part I exam.

6. **Is an emergency plan established in case there are problems at the test site such as noise, fire, flood, or other interruptions?**

Yes.

7. **What is the procedure for providing feedback to the credentialing authority regarding irregularities in administration, such as cheating?**

Thomson Prometric keeps the NBPME informed of irregularities occurring during the administration and scoring of the Part I exam that would affect the validity of test scores. If necessary, actions consistent with industry practices are taken to maintain the validity of test scores.

- Are irregularities evaluated for impact on validity of test scores before scores are released to candidates and others?

Yes.

- Is there a policy for withholding scores when validity is at issue due to an irregularity?

Yes.

7. Is the test equated so that reported scores on different forms will be a result of candidate ability and not item difficulty?

No. See answer to question #3 under the **AUDIT OF STATISTICAL ANALYSES AND SCORING** section of this report.

- What method is used to equate the tests?

Not applicable.

- Why was that method selected?

Not applicable.

- On what sample group was the equating performed?

Not applicable.

9. **What are the pass/fail rates, and the sample groups on which they were calculated? Are the pass/fail rates reviewed for expected outcomes, given the sample group?**

See response to question #4 under the **AUDIT OF STATISTICAL ANALYSES AND SCORING** section of this report.

10. **If the test is comprised of more than one scored part, are the scores combined into one total score, or must the parts be passed separately?**

The test consists of items classified according to seven basic science content areas. Candidate performance on these items is summed to determine a candidate's total score on the Part I exam. Pass/fail status is based on a candidate's total score.

- Is there a study or other documentation that supports the decision whether to combine scores or keep them separate?

Yes.

- If scores are combined, what is the method of combining them?

A simple summation process is used.

- If scores are separate, are they each sufficiently reliable? Do the parts represent separate content?

Not applicable.