



Standard Setting Report for TMOD*
February 2010

Background

As part of the NBEO restructure, the Part II Patient Assessment and Management (PAM) exam now contains the Treatment and Management of Ocular Disease (TMOD) items. Of the 350 items on the PAM exam, approximately 100 – 120 are categorized as TMOD items. Since a TMOD breakout score and pass-fail decision will be reported for state board requirements, a passing standard needed to be established. The purpose of this report is to summarize the standard setting study conducted to establish the cut score for TMOD.

In order to be considered part of TMOD and to be counted within the breakout score, the content of an item must pertain to one or more of the following:

- Formulation of most appropriate disease diagnosis which will be treated and/or managed
- Clinical correlation of basic science principles related to disease diagnosis and treatment
- Selection of treatment/management, including systemic considerations
- Dose, form, schedule, and duration of treatment
- Contraindications and side effects of medication, including systemic considerations
- Follow-up and prognosis, including reassessment of diagnosis after initiating treatment
- Treatment and management of ocular emergencies and urgencies

Because of the case structure of the Part II-PAM exam, the classification of TMOD was slightly expanded to include items pertaining to disease diagnosis which will be treated and managed with a therapeutic agent. The discrete item format of the Part II Clinical Science exam allowed for a diagnosis to be built into the item stem and permitted the extraction of a score based on items that exclusively dealt with treatment and management. Part II-PAM requires the candidate to consider all data presented in the case and to make a diagnosis before proceeding to treatment. Diagnosis is an integral part of the task of the clinician; thus these items are included in the TMOD score.

Some items classified as TMOD for Part II-PAM are also considered “Clinical Correlation of Basic Science.” These items were integrated into PAM from the Clinical Science exam, where they were classified as TMOD.

* The National Board wishes to acknowledge the efforts of the many individuals who have contributed to the development of the Part II-PAM examination. It would have not been possible without the generosity and expertise of the many case authors, the Part II-PAM Committee and Council members.

It should be noted that the TMOD items were rated by panelists in the context of all items in the case and candidates will respond to these items under the same conditions. For that reason, any potential cuing from other non-TMOD items within the case has been accounted for by the standard setting method.

Procedure

Method

The cut score for TMOD was established using the data collected from the Part II-PAM standard setting study, so the method is described in the context of the Part II examination. The complete Part II-PAM standard setting report presents an overview of various standard setting techniques and contains many more details of the study.

The PAM standard setting study was conducted using the Yes/No method (Impara & Plake, 1997), a method based on the Angoff (1971) approach. This procedure is based on the notion that if it were possible to positively identify a large group of marginal candidates, they could be given the test and their test scores analyzed. Since this group has already been identified as “borderline,” it should be acceptable that some of this group would exceed the passing standard, while others would not. As a result, it would be reasonable to establish a “cut score” at their average score.

Since it is not practical to identify borderline candidates, give them a test, and then compute their average score, another plan must be used. This plan involves three steps: (a) asking a group of experts to identify a hypothetical marginally competent candidate, (b) asking them to decide whether or not that candidate would get each item correct; and (c) computing the average scores over all judges and over all of the items. The sum of the item rating for a given judge is an expected score for a minimally qualified examinee. By asking the judges to make their determinations independently, it is reasonable to average their scores to derive a passing standard that represents, in some sense, a consensus agreement on the expected performance of minimally qualified candidates.

The choice of the Yes/No method for the Part II examination was based on a number of factors:

- The structure of the PAM exam, which is case-based, means that there is a certain amount of interdependence among the items associated with each scenario. A simple yes/no decision for each item allowed the panelists to take the relationships among the items into account as they made their judgments.
- The examination is very long and comprehensive with 60 cases and 350 items. In order to increase the probability that the judges would be able to complete the task of rating all of the items, a simple cognitive task (yes/no) was desirable.
- In the past, the NBEO has made extensive use of the Nedelsky method and many of the panelists in this study have participated in its use. The yes/no method offered the convenience of being an item/test centered technique and at the same time, it was a direct contrast with other methods that involve estimating item difficulties (e.g., Nedelsky, Angoff).
- Because the Part II-PAM exam has not yet been administered, it was not feasible to use an ordered item booklet, such as with the bookmark method.

Orientation

Each of the meetings ran for eight hours on August 5, October 5, and October 24 respectively. After a general orientation, participants discussed different approaches to standard setting and evaluated the advantages and disadvantages of both domain-referenced and norm-referenced standards. In addition, the various methods of setting domain-referenced standards were discussed along with a rationale for the use of the Yes/No method.

Approximately 3 hours was spent in training and orientation before the actual standard setting exercise. The primary goals of the the training was to:

- Provide an overview of the PAM exam, in particular the case structure and the types of items within each case. The ABS groups were provided with an extensive overview of the Part II-PAM examination and a comparison with the original Part III-PAM.
- provide an overview to standard setting and to convey a sense of the importance of the process,
- Discuss case/item difficulty and to teach judges to estimate the difficulty of item. Several cases from the April 2008 Part III-PAM exam were used to facilitate the training process.
- Introduce the concept of minimal competence and to help judges arrive at their own personal definitions of borderline performance.

During the discussion of case/item difficulty, panelists were directed: *For each of the items in the case, indicate with 'Y' or 'N' whether or NOT a minimally proficient candidate **would** get the item correct.* Judgments were recorded, averaged, and then compared to the actual examinee performance. Participants discussed their individual judgments and shared the reasoning behind their decisions.

The discussion of minimal competence (borderline performance) is an important part in the implementation of the Yes/No method. This conversation also focused on a group definition/perception of what constitutes 'entry-level.' To facilitate the discussion the marginally qualified candidate was contrasted with both expert and incompetent candidates. Expert candidates were characterized as having integrated knowledge; they are able to make connections between abstract concepts and concrete patient situations. In addition, experts have a wider breadth of knowledge which is often at a deeper level. The incompetent/unqualified candidate was characterized as someone who has a minimal amount of knowledge, is often unaware or unwilling to acknowledge their deficiencies and as a result, appears arrogant.

The marginally qualified candidate was characterized as someone with a 'thin' base of knowledge who wants to be told exactly how much they need to know. In addition, this individual knows just enough about the range of topics required in entry level optometry. This level of knowledge was characterized as 'bits and pieces' in one standard setting group and as 'fragmented' in another. The marginally qualified candidate does not have an integrated knowledge of optometry. In addition, he or she may not be a very curious person. Finally, the minimally qualified optometrist could be a very practical person who might have struggled to master basic concepts and who recognizes their limitations.

Standard Setting Judgments

After panelists completed training, the December 2009 Part II-PAM examination was distributed. Participants also received the exam key (correct answers), and a form for recording their judgments. In order to account for fatigue effects among raters, panelists were asked to begin the case/item rating process at different points among the 60 cases. In other words, everyone did not start at Item 1 and proceed to Item 350.

The directions to panelists for the standard setting ratings were identical to the ones used in the training process:

*For each of the items in the case, indicate with 'Y' or 'N' whether or NOT a minimally proficient candidate **would** get the item correct.*

After the rating process, judgments were collected, entered into a spreadsheet, tabulated and summarized. Participants were not given the results of the tabulation and were not offered an opportunity to revise their estimates, as is common in some standard setting studies.

Results

Table 1 presents the individual judges' ratings for the 111 Part II-PAM TMOD items. These ratings were obtained during the standard setting exercise and were extracted for this analysis. Data from 23 panelists were included in the analysis; one participant's ratings were excluded because they were incomplete.

Table 1. Individual Judge Ratings for TMOD items

<u>Judge</u>	<u>TMOD items</u>
1	0.595
2	0.523
3	0.631
4	0.667
5	0.622
6	0.631
7	0.748
8	0.604
9	0.721
10	0.712
11	0.712
12	0.748
13	0.586
14	0.775
15	0.712
16	0.613
17	0.739
18	0.910
19	0.686
20	0.793
21	0.725
22	0.820
23	0.640
Mean	0.692

The mean of the item ratings for the TMOD items resulted in a cut score of 0.692 or 69.2% correct. The TMOD cut score (69.2%) is slightly higher than the overall Part II cut score (68.1%). This indicates that the panelists felt that these items were slightly easier than the total exam. The NBEO Board of Directors adopted this cut score for the TMOD section of the December 2009 Part II exam. Candidates correctly responding to 77 (or more) out of 111 items received passing scores for this section of the exam.

To maintain this passing standard on future administrations, the Rasch item response theory model (1960) will be used to equate sets of TMOD.

References

- Angoff, W.H. (1971). Scales, norms, and equivalent scores. In R.L. Thorndike (Ed.) *Educational measurement* (pp. 508-600). Washington, DC: American Council on Education.
- Impara, J.C. & Plake, B.S. (1997). Standard setting: An alternative approach. *Journal of Educational Measurement*, 34, 353-366.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.