

How The National Board Establishes Its Pass-Fail Standards on Multiple-choice Tests*

One of the most important characteristics of the National Board examinations is criterion-referenced pass-fail standard setting. Since this procedure does not require a predetermined failure rate, candidates do not compete against each other for successful scores. Instead, they compete against a standard, analogous to the criterion of par in golf. When implemented in 1981, NBEO became the first national board for a licensed profession to adopt this type of procedure. The manner in which pass-fail cutoff scores are determined is described in this article.

Methodology

The standard setting procedure that the Board uses is one of the variations of the Nedelsky technique, named after the science educator who created it. The underlying philosophy of this technique is that the difficulty of a test item, and hence, the level of examinee performance that should be expected, is largely dependent on the degree to which the incorrect (i.e., distractor) options approximate the correct response. For example, consider the following items.

More Difficult	Less Difficult
Which of the following cities is the capital of Illinois?	Which of the following cities is the capital of Illinois?
a. Chicago b. Springfield	a. Rockford b. Springfield
c. Urbana d. Peoria	c. Urbana d. Peoria

The question or stem portion of each item is identical. However, the first version is more difficult because distractor option A (Chicago) is more compelling than its counterpart response (Rockford) in the second version of the item. An examining board would therefore expect a higher level of performance on the second item.

The Nedelsky technique establishes an option rating scale to quantify the level of performance that should be expected as a function of the distractor options. The values of the rating scale are presented below.

2 = the one best or correct response

1 = a sophisticated or plausible distractor which the examinee would have difficulty distinguishing the one best or correct response.

0 = an unsophisticated or implausible distractor which the examinee would be expected to eliminate.

This scale is used for standard setting only. Scoring is conducted in the usual dichotomous manner, with zero points awarded for selecting a distractor, regardless of its standard setting weight, and one point awarded for the best or correct response. From the rating scale, an item standard, known as a minimum pass index (MPI), is computed using the following formula:

$$\text{MPI} = \frac{2}{\text{Sum of weights}} - \frac{1}{5 (\text{Sum of weights})}$$

If Chicago were regarded as a sophisticated distractor in the first example, the item MPI would be computed in the following manner:

$$\text{MPI} = \frac{2}{\text{Sum of weights}} - \frac{1}{5 (\text{Sum of weights})}$$

$$\frac{2}{1+2+0+0} - \frac{1}{5(1+2+0+0)}$$

$$= (.67 - .07) = .60$$

This MPI, in essence, establishes the level of expected performance or "par" for this item. The pass-fail cutoff score is then based on the average of the item MPIs. If, for example, the MPI that was just computed was for a typical item on a test, candidates would need to correctly answer 60% of the items in order to pass.

Since the values that the item MPI can have are a function of the number of sophisticated but incorrect options, the array of MPI values is limited, as shown in the table below.

Range of MPI Values

Number of Sophisticated but Incorrect Options	MPI
0	.90
1	.60
2	.45
3	.36
4	.30

Implications

The MPI weights are assigned consensually by the examination committee members during the construction of a particular test. To illustrate the magnitude of this task, consider Part I of the National Board, consisting of 435 items. The test is developed by four committees; one per section. Nearly 20 subject matter experts (including faculty and independent practitioners) are involved in setting the Part I passing standard. Since the Part I items currently contain an average of 4 options per item (i.e., 3 distractors per item), there are approximately 1,300 consensus judgmental entities that comprise the passing standard.

With constantly changing candidate characteristics resulting from new academic institutions, a fluctuating number of applications, changing candidate backgrounds, and changing licensure statutes, norm-referencing (i.e., grading on a curve) has little justification. There is substantial pressure being exerted on all of the national boards to adopt criterion-referencing; the profession of optometry should be proud that its national board was the first among the licensed doctoral-level professions to implement this standard setting approach.

* Revised from Gross, L.J., *The Examiner*, February-March 1981