

Equating Tests and Use of Scaled Scores

Question: *We use a criterion-referenced methodology when establishing the passing score for our examination. The result from applying this methodology is that the actual passing score varies from examination to examination. What is the best way to explain the variation?*

Answer: The source of the variation is easy to explain. Since the difficulty of the questions selected for a given administration of an examination differ from those of another administration, the passing score varies around the fixed concept of minimal competence to account for the differences in the examination questions. That is, the examination with the more difficult questions overall will have a lower passing score. The variation in the passing score is psychometrically sound and legally defensible. As you mentioned, the variation is difficult to explain to candidates who may achieve a score that fails them on one examination but would be a passing score on a subsequent examination.

A good way to reduce or even eliminate varying passing scores is to assemble tests that are equal in difficulty according to average difficulty (p values). The passing score is not the equivalent of the average of the p values; however, if the average difficulty of the examinations were about the same, then the passing scores should not vary significantly. The limitation for this method is that the items should have stable statistics. The stability usually is not obtained until the items have been administered to several hundred candidates. For small examination programs, stable item statistics may not be available.

To avoid this situation, the best option is to report the results by scaled scores. When a scaled criterion-referenced passing score is used, the score required for passing remains consistent. Scaling the score does not affect the level of performance required for passing the examination. Scaling simply allows the licensing board to report, for example, that the passing score is 70 (not to be confused with 70%) while the actual passing score is free to vary according to the difficulty of the examination.

Note that it is easier to construct equivalent forms if new items are pretested, e.g. included on a previous form as unscored items.

After they have been vetted with statistics, the next test form will have statistics for all scored items, and there will be little need to modify the percent correct passing score (or derived scaled score) based on performance of a new group of candidates.