

## **ACGME'S OVERVIEW OF STANDARDS FOR EVALUATING THE QUALITY OF ASSESSMENT METHODS**

### Reliability

1. Reliability indicators must be available for any total score or subscore that will be interpreted
2. Interrater and intrarater reliability for multiple ratings of the same learner should be proved when scoring or rating entails subjective judgment
3. For high-stakes decision, an estimate should be proved of the percentage of learners who would be classified the same on 2 applications of the same method or rating process

### Validity

1. A rationale for each interpretation and use of evaluation results along with evidence and theory should be presented.
2. Processes and procedures used for selection of the content of assessment and for any criteria (e.g., importance, frequency, and criticality) used to sample content should be described and justified when validation rests in part on the assessment content.
3. When the rationale for the use and interpretation of an assessment depends on the psychological processes or cognitive operations of the learner or the processes of the evaluator, the theoretical or empirical evidence that supports the interpretation should be provided.
4. When unintended consequences result from use of a specific assessment, an attempt should be made to identify the cause. For example: Is the assessment measuring something other than what it was intended to measure? Did the assessment fail to measure fully the intended construct?
5. The degree of agreement between a single expert rater and "gold standard" or consensus ratings for the same performance should be provided when a single rater using subjective judgments is the basis of the assessment.
6. When a single rater using subjective judgment is the basis of the assessment, the degree to which known strengths and weaknesses of the learner are detected should be provided.

### Ease of use

1. The assessment tool is easily carried or accessed in the course of daily clinical or teaching activity.
2. The tool requires little special setup.
3. The tool requires less than 20 minutes for the assessor to complete.

### Resources required

1. No additional resources are required beyond the documentation tools.
2. Training requirements for assessors do not exceed an hour.
3. No additional persons other than an individual assessor are required to complete the evaluation.

### Ease of interpretation

1. Individual scores are interpretable-for example, on an easily understood scale, such as percent correct or against behavioral or other descriptive criteria-and are accompanied by interpretation guidelines.

2. Normative data are available consisting of: (1) a standard of care; (2) performance of other residents at the same level of training and/or experience; (3) performance of other residents with more or less experience; and (4) the resident's performance level at an earlier stage of education and experience.
3. Preprogrammed, easy-to-read reports and graphs make it simple to compare individual to group performance.

#### Educational impact

1. The method has been shown to positively affect individual learner performance; that is, there is a change in knowledge, skills, or attitudes.
2. The method has been shown to positively affect or change program curriculum (should be corroborated in at least 2 studies)
3. The method has been shown to provide specific actionable results that are regarded as useful by the learners.

#### Grading for educational impact

- A. Meets both standard 1 and 2.
- B. Meets either standard 1 or 2.
- C. Standard 3 is met.
- NI Not enough information from literature to judge

#### Standard for educational impact

1. The method has been shown to positively affect individual learner performance (change in knowledge, skills, or attitudes).
2. The method has been shown to positively affect or change program curriculum (should be corroborated in at least 2 studies).
3. The method has been shown to provide specific actionable results that are regarded as useful by the learners.