KEY DEFINITIONS in Assessment

SUMMATIVE VS FORMATIVE

Summative assessments are designed to assess skills and make a classification decision about an examinee *after* instruction is completed. ABIM summative assessments measure the skills necessary for providing safe, effective healthcare in an unsupervised setting; the scores are used to make a high-stakes pass/fail decision about physicians.

Formative assessments are designed to provide feedback and guide learning, and are administered *before* instruction begins or *during* instruction. These assessments are typically used in educational contexts where informing learning and instruction is the primary focus, and high-stakes decisions are typically not the intention.

PASSING SCORE VS PASS RATE

A **passing score** is the number of questions an examinee must get correct to pass a test. For example, a passing score of 60 means that an examinee must answer at least 60 questions correct to pass. Once established, the passing score is constant for many years.

A **pass rate** is the percent of examinees who pass the test. For example, a 90% pass rate means that 90% of examinees passed the test in a given administration. The pass rate for a test can change each year due to differences in the ability of the examinees.

STANDARDIZED TEST VS CLASSROOM TEST

A **standardized test** is one in which the content, format, and test administration procedures are controlled such that they are consistent and fair for all examinees. Test questions are evaluated to ensure that they are of appropriate difficulty, relevance, and free of bias. Test scores from standardized tests have consistent interpretation across test forms and years.

A **classroom test** is typically a small-scale test administered without consistent procedures or standardized, controlled attention to fairness. Test questions are *not* typically reviewed or evaluated to ensure that they are of appropriate difficulty, relevance, or are free of bias. Test scores from classroom tests do not typically have consistent interpretation across forms and years.

RELIABILITY VS VALIDITY

Reliability is the degree to which test scores are consistent and reproducible. In other words, it quantifies the degree to which a test consistently returns the same score when repeated under similar conditions. For example, if one repeatedly steps on the same scale to determine weight, a scale that consistently gives the same weight is giving a reliable number. Reliability is a necessary but not sufficient property of validity.

Validity is the degree to which test scores are appropriate, meaningful, and useful for making specific inferences and decisions; it is the degree to which a test score accurately depicts what it purports to measure. For example, if one steps on a scale and it reads 3 pounds, the scale is not providing a valid weight. Valid test scores require a chain of evidence that connects scores to some outcome(s) of interest.

American Board of Internal Medicine®

510 Walnut Street, Suite 1700, Philadelphia, PA 19106