



**Glossary of Testing Terminology  
Commonly Used in Professional and Occupational Regulation**

---

This document provides definitions for words and phrases commonly used in the field of professional and occupational regulation. By design it is limited in scope and aims to assist meaningful conversations between international regulators. This project was initiated by the International Relations Committee of the Council on Licensure, Enforcement and Regulation (CLEAR), whose members include regulators from the United States, Canada, Mexico, and the United Kingdom. Additional resources and further information about CLEAR can be located at <http://www.clearhq.org/>

Please note: this is a living document and comments and suggestions are welcomed. Please contact [aparfitt@mis.net](mailto:aparfitt@mis.net) with your ideas or to submit a new or revised definition.

<b>Accession number</b>	An alphanumeric unique identifier for a test item typically including a program code and a serialized ID.
<b>Adverse impact</b>	The disproportionately negative results that a law, process or policy may have on a specific group(s) of individuals that share certain traits, characteristics or other discerning features, which may include race, culture, gender, or other non-relevant factors.
<b>Anchor items</b> <i>(see also Equating set)</i>	Items used on both a new test form and an administered test form for the purpose of equating the difficulty level of the two tests.
<b>Angoff Method of Standard Setting/Modified Angoff</b>	One of several procedures for establishing a threshold for passing the examination that is tied to successful performance on the job for an entry-level or minimally competent candidate, or for a level of certification beyond entry to practice (e.g. advanced or specialty practice). In Angoff procedures, a group of subject matter experts makes judgments about the difficulty of each question on a criterion

referenced test. The judgments are combined mathematically to arrive at a recommended minimum score (see Passing Point).

**Candidate handbook**

A booklet, including examination information, for those applying for a credential, e.g., certification, licensure, registration.

**Common item equating**

A process to ensure comparable scores when common items appear on two tests. Also the basis for establishing the difficulty level and passing score for the new test form. (See Equating.)

**Computer adaptive testing (CAT)**

A non-linear computer-based examination in which each successive item is based on a candidate's performance on the previous item. The procedures may result in a unique examination for each candidate and may be of fixed or variable length.

**Computer mastery testing**

A computerized test on which candidates must obtain an established mastery score on each section of the test.

**Constructed response**

A type of item in which the candidate produces a response rather than just selecting a response from a set of multiple-choice options.

**Content classifications/Content outline/Test blueprint**

A scheme for classifying the terms in an item bank based on the content measured by those items. The scheme should specify the number of questions on the test that are drawn from each content area or topic to ensure that the content covered by each test form is consistent. The scheme is usually based on the results of a job analysis in order to provide a link between the practice of the profession and the content covered on the test. This link provides evidence for the validity of the test.

**Correlation coefficient**

A measurement of performance on a particular test item compared to performance on the test as a whole. A

strong positive correlation means the high ability candidates are getting the item right. A strong negative correlation means the high ability candidates are getting the item wrong. A near zero correlation means the item is not discriminating between high and low ability candidates. Typical correlation coefficients used in testing include biserial correlation coefficient and point-biserial correlation coefficient.

**Criterion-referenced testing**

Testing in which a fixed passing or cut score is set using accepted standard-setting methods (compare Norm-referenced testing.)

**Cut score** (*see Passing point*)

**Defensibility**

A test is considered defensible if proper psychometric procedures are followed when establishing the content and statistical specifications and when developing and scoring the test and when setting standards or making decisions on the basis of test results. A defensible test must demonstrate both reliability and validity. In the legal context, this term can be defined as "the extent to which a position can be defended."

**Diagnostic score report**

A summary of a candidate's performance, both positive and negative, on subsections of a test, usually intended to provide information on relative areas of strength and weakness. Generally only provided to failing candidates.

**Difficulty level**

The percentage of candidates who answered a specific item correctly or the average scale value or rating achieved by candidates. Also called a P Value.

**Distracters/Distractors**

The incorrect options in a multiple-choice item.

**Equating**

Statistically aligning reported scores on a test form to ensure comparability of scores across test forms in a given

	testing program.
<b>Equating set</b> <i>(see also Anchor items)</i>	A group of items that are the same in a new test form as in the original test form.
<b>Evaluation instrument</b> <i>(see Examination)</i>	
<b>Examination</b>	An evaluation instrument that measures a candidate's competency by one or more means, such as written, oral, practical, or observational (ISO/IEC 17024).
<b>Examiner</b>	A person with relevant technical and personal qualifications who conducts and/or scores an examination (ISO/IES 17024).
<b>Face validity</b>	The degree to which a test appears to be an appropriate measure of the knowledge, skills, and abilities being tested.
<b>Fairness review</b>	A process used to ensure that examinations or other processes do NOT contain any items that are potentially discriminatory or insensitive toward a particular group of examinees.
<b>Field-test item</b> <i>(see Pretest item)</i>	
<b>Fixed-length test</b>	A test with a preset number of items in each form of the examination.
<b>Form code</b>	A unique identifier for a test form that typically includes the testing program code and a series of alpha-numeric characters.
<b>Inter-rater reliability</b>	The degree of consistency with which different raters assign scores to candidates' performance.
<b>Interstate score transfer</b>	A test score transferred directly to a jurisdiction by a non-state agency for a candidate who is seeking licensure or registration in a new jurisdiction.

<b>Interview test/Oral examination</b>	A testing method in which a predetermined set of items is orally presented to all candidates. Responses are scored with a rating guide designed by a committee of subject matter experts. Examiners are carefully trained to ensure comparability of ratings, and typically more than one examiner rates each candidate. Candidates are typically videotaped or audio taped in order to have a record of the items asked and responses provided in the event that a candidate appeals the results.
<b>Item</b>	A generic term indicating a single point of measurement in an examination; a test question or other test unit such as a performance task.
<b>Item analysis</b>	A report of the difficulty and discrimination for each item on an examination. The analysis provides statistical information about the correct response and each distracter/distractor.
<b>Item bank</b>	A repository, generally in electronic format, for all of the items associated with a particular testing program. In addition to the questions, item banks normally contain content classification information and statistical information.
<b>Job analysis/Occupational analysis/ Practice analysis</b>	A process for describing the practice of a profession or job, including underlying competencies, major areas of responsibility, tasks, and knowledge, skills, and abilities. Results can be used to determine the content that should be covered on the examination.
<b>Key</b>	The correct answer to an individual test question.
<b>KSAs</b>	Knowledge, skills, and abilities associated with the practice of a profession or occupation.
<b>Linear test</b>	A test in which a specific set of items is administered to all candidates without taking into account a candidate's ability

level.

**Linear-on-the-fly test**

A computer based test in which varying tests are assembled for candidates based on content and statistical test specifications. This format differs from adaptive testing in that the difficulty of test items is not based on the candidate's performance on earlier items in the test.

**Mean**

The arithmetic average of a set of numerical data. In a testing context, it refers to the "average" score obtained by a group of candidates.

**Median**

The middle value of an ordered set of numerical data. For example, the median value of the set {5, 8, 9, 10, 11, 11, 13} is 10.

**Minimum requirements**

The thresholds that must be met for eligibility to sit for an examination, to pass the examination, or to be qualified for a credential.

**Mode**

The most frequently occurring value in a set of data. For example, the mode of the set {13, 5, 9, 11, 11, 8, 10} is 11.

**Modified Angoff** (*see Angoff method of standard setting*)

**Multiple choice question**

An item that lists multiple response options, one of which is clearly the correct or best option.

**Norm-referenced testing**

Testing in which candidates' scores are reported in relation to the performance of the overall group taking the test. The test scores may be reported as percentile ranks or scaled scores, e.g., 200 – 800 or similar scale (compare Criterion-referenced testing).

**Obtained score**

The score a candidate earns on a test, which includes measurement error (see True score and Standard error of measurement).

**Occupational analysis** (*see Job analysis*)

**Operational item**

An item on a test that is scored and contributes to the pass/fail decision (see Pretest item).

**Options**

The various responses in a selected-response test question from which a candidate would select the correct answer (see Selected response).

**Oral examination** (*see Interview test*)

**P-value**

Percentage of candidates answering an item correctly, or the average rating achieved by candidates on a scale of 0 to 1. The P-value may range from 0 to 1.0 (see Difficulty).

**Passing point/Pass mark/Passing score/Cut score**

The score (mark) required to pass an examination or achieve a particular result/classification.

**Percentile**

A value on a scale that indicates the percent of a distribution that is equal to it or below. For example, a score at the 95th percentile is equal to or better than 95 percent of the scores.

**Pilot item** (*see Pretest item*)

**Practical examination**

A performance-based test based on requirements of a job or the standards of practice of a profession; a measure of an individual's skill.

**Practice analysis** (*see Job analysis, also Standard setting*)

**Pretest item/Pilot item/Field test item**

Items included in an examination solely for the purpose of collecting statistical data. The items do not count towards a candidate's score.

**Prior Learning Assessment and Recognition**

The process of identifying, assessing and recognizing skills, knowledge, or competencies that have been acquired through work experience, previous education, independent study and other

activities. Prior learning may be applied toward academic credit, requirement for entry to practice to an education/training program or for certification.

**Psychometrics**

The field of study connected to psychology and statistics concerned with the measurement of psychological aspects of a person such as knowledge, skills, and abilities.

**Raw score**

The total number of operational items answered correctly on a test; the sum or mean of all ratings achieved on a performance test.

**Reference list**

Source materials for the content of the test as well as a list supplied to candidates to prepare for the test.

**Reliability**

The degree to which a test consistently measures performance, e.g., within items, across occasions, across raters.

**Scaled score**

The conversion of a raw score (i.e. number of correct responses) to a special scale used for reporting purposes. Commonly used on equated test forms so that the passing score remains constant even though answering different numbers of items correctly may be required. For example, many tests report scores on a 200-800 score scale.

**Selected response**

A type of item in which candidates must choose from options presented, e.g., multiple choice, matching, drag and drop.

**Self-assessment**

Voluntary measurement of one's knowledge, skills, and abilities in a certain area. May be considered to be a low-stakes test in that the results do not impact licensure status.

**Special accommodations**

Special testing conditions provided for people with disabilities, e.g., Braille form, additional time, separate testing rooms, etc. For example, the Americans

with Disabilities Act (ADA) is a 1990 U.S. Federal law that prohibits discrimination against individuals with disabilities by public or private entities in the areas of employment, public accommodations, state and local government services, telecommunication, and standardized testing. In Canada it is Human Rights legislation in each province. The United Kingdom (UK) equivalent is the Disability Discrimination Act 1995.

**Standard deviation**

A measure of the variability of a distribution of scores. The more the scores cluster around the mean, the smaller the standard deviation. In a normal distribution, 68% of the scores fall within one standard deviation above and one standard deviation below the mean. (Square root of the variance).

**Standard error of measurement**

An estimate of the measurement 'error' associated with the test-takers' obtained scores when compared with their hypothetical 'true' scores. The amount of variation that is expected in a candidate's test score if the candidate were able to take a test many times (without a change in the knowledge level). The calculation is based on the reliability of the test and the standard deviation of the score distribution. (Sometimes called an error band).

**Standard setting** (*also Practice standards, or Standards of practice*)

The process used to establish the minimum score needed to pass an examination. Many methods exist, ranging from regulatory bodies that arbitrarily select a score (e.g. 70 percent correct) to formal processes based on the collective judgement of a group of subject matter experts. In order to be defensible, passing scores should be established using psychometrically sound procedures.

**Standards of practice** (*see Standard setting*)

**Stem**

The premise, including the facts/details, around which an item is structured; the

portion of the item that poses the question or presents the problem.

**Test adaptation**

Adapting a test for use in other languages/cultures, typically referring to the process of translating a test from the source language to a target language. However, true adaptation goes beyond a literal translation of the test and will include changes based on cultural differences.

**Test blueprint** (*see Content classifications*)

**Test specifications**

The content outline, test blueprint, and statistical requirements for a specific testing program.

**True score**

The score that a candidate would obtain on an examination in the absence of measurement error. This theoretical score represents the exact amount of knowledge that the candidate possesses.

**Validity**

The degree to which a test measures the content it purports to measure. Validity evidence may be content-based, construct related or predictive. In criterion referenced testing, content based evidence from a job analysis is generally considered to be critical.

**Variable-length test**

A computer adaptive test that varies in the number of items administered. The test concludes when enough information has been collected to establish the ability level of the candidate.

**Last revised: August 27, 2007**