



**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. Recent revisions were made under the direction of the Examination Resources and Advisory Committee. 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 <http://www.clearhq.org/> with your ideas.

Ability/Trait	The underlying knowledge, skill or proficiency of a candidate that is assessed with a test.
Accession number	An alphanumeric unique identifier for a test item that stays with the item throughout the life of the item.
Adverse impact	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, but is not limited to, race, culture, gender, or other non-relevant factors.
Anchor items (see also Equating set)	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.

Angoff Method of Standard Setting/Modified Angoff

One of several frequently used procedures for establishing a threshold for passing the examination that is associated with 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 the modified Angoff procedure, a group of subject matter experts estimate the percentage of minimally competent candidates who will correctly answer each item on the test. The estimates are summarized mathematically to provide decision makers with a set of scores from which a recommended minimum score is chosen (see Passing Point).

Candidate handbook

Information about a test and/or a testing program that is provided to a candidate applying for a credential, e.g., certification, licensure, registration.

Classical test theory

A psychometric theory based on the view that an individual's observed score on a test is the sum of a true score component plus an independent measurement error component.

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, or sets of items, is based on a candidate's performance on the previous item(s). Using CAT may result in a unique examination for each candidate that may be of fixed or variable length.

Computer Based Testing

Examination administered via a computer.

Computer mastery testing

Computerized tests on which candidates must obtain an established mastery score on each section of the test. It

	may be of fixed or variable length.
Constructed response	A type of item in which the candidate produces a non-structured response (e.g., essay, oral exam) rather than selecting a response from a set of options.
Content classifications/Content outline/Test blueprint	The content outline specifies the number of questions on the test that are selected from each content area or topic to ensure that the content covered on each test form is consistent. The content outline is based on the findings from a job analysis which provides a link between the practice of the profession and the content covered on the test. This link provides evidence in support of the validity of the test.
Correlation coefficient	A statistical value that measures the strength and direction of the relationship between two variables.
Criterion-referenced testing	Testing in which a fixed passing or cut score is set using accepted standard-setting methods and is used to evaluate all candidates (see 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, 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.
Diagnostic score report/ grade 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 required for failing candidates.
Difficulty level/Facility index	The proportion of candidates who

answered a specific item correctly or the average scale value or rating achieved by candidates. Also called a p -value.

Discrimination Index

The extent to which performance on a particular test item compares 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.

Distracters/Distractors

The incorrect options in a multiple-choice item.

Equating

The construction of two or more test forms covering the same content, which conform to similar statistical specifications, are administered under identical procedures, and are based on a common score scale.

Equating set (see also Anchor items)

A group of items that are the same in a new test form as in the original test form used to establish a link between test forms.

Evaluation instrument (see Examination/Test)

Examination [Test]

An evaluation instrument used to make inferences about a candidate's competency by one or more means, such as written, oral, practical, or observational (ISO/IEC 17024).

Examiner

A person competent to conduct and score an examination, where the examination requires professional judgment (ISO/IES 17024).

Face validity

The degree to which a test appears to be an appropriate measure of the knowledge, skills, and abilities being

Fairness review

evaluated.

A process used to ensure that examinations or other processes do NOT contain any items that are potentially discriminatory or biased toward any particular group of examinees.

Field-test item (see Pretest item, pilot item)

Fixed-length test

A test with a preset number of items in each form of the examination.

Form code/test code

A unique identifier for a test form that typically includes the testing program code and a series of alphanumeric characters.

Inter-rater (Inter-examiner) reliability

The degree of consistency with which different raters are in agreement with each other, such as when they assign scores to candidates' performance.

Interstate score transfer

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.

Interview test/Oral examination

A testing method in which a standardized set of items is orally presented to all candidates. Responses are scored with a standardized rating guide designed by a committee of subject panel 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 to provide a record of the examination.

Invigilator/Proctor/Supervisor/Test Administrator

A person authorized by the certification body who administers or supervises an examination but does not evaluate the competence of the candidate (ISO/IEC 17024).

Item/question

A generic term indicating a single point

of measurement in an examination; a test question or other test unit such as a performance task.

Item analysis

A report of the difficulty and discrimination for each item on an examination. The analysis provides statistical information about the correct response and may include information about each distracter/distractor.

Item bank

A repository, generally in electronic format, for all of the items associated with a particular test or testing program. In addition to the questions, item banks may contain associated information such as content classification, statistical information, reference citations, item author, and a log of changes made to an item during its use.

Item response theory

A theory of measurement that attempts to make stronger assumptions than classical test theory makes. Also known as latent trait theory, IRT requires complex mathematical models that assume a single underlying trait for examinee ability. The probability of a correct answer relates to the candidate's level of that ability. Typically, up to three parameters can be computed; an S-shaped item characteristic curve (ICC) usually represents the probability of a correct response. A common model only considers item difficulty, and is known as the Rasch or 1-PL model.

**Job analysis/Occupational analysis/
Practice analysis**

A process for describing the practice of a profession, occupation, or job, including underlying competencies, major areas of responsibility, tasks, and/or knowledge, skills, and abilities. Results can be used to determine the content of an examination. Items developed for an examination are classified according to the job analysis. A job analysis provides evidence in support of the legal defensibility for a test.

Key/answer

The correct response to an individual

test item.

KSAs

Acronym for knowledge, skills, and abilities -- these terms are associated with the practice of a profession, occupation, or job.

Linear test

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 (LOFT)

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 qualify to take 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, single option

An item that lists multiple response options, of which only one is 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. Generally not used for credentialing examinations.

Obtained/observed 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, Practice analysis)

Operational item

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

Options

The various responses in a test question from which a candidate would select the correct answer. The correct option is called the key; incorrect options are called distracters/distractors.

Oral examination (see Interview test)

***p*-value**

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

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 Field test item, 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. The same psychometric requirements apply as for

a selected-response (e.g., multiple-choice) test.

Practice analysis (see Job analysis, Occupational analysis)

Pretest item (see Field test item, Pilot item)

Items included in an examination solely for the purpose of collecting statistical data. The items do not count towards a candidate's score. Candidates are generally unaware which items are pretest items and which are scored items.

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 or may be a 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.

Rasch model (see IRT)

A psychometric item response theory (IRT) model for analysing test scores. It is the simplest of such models.

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 different scale used for reporting

purposes. Commonly used on equated test forms so that the reported passing score remains constant even though answering different numbers of items correctly may be required.

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.

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.

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 candidates 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 for conducting a passing score exist. 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 or localization

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. Process of modifying an examination for use in other languages or cultures based on linguistic, cultural and practice pattern differences.

Test blueprint (see Content outline)

Testing accommodations

Alternate testing accommodations provided for people with disabilities, e.g., Braille form, additional time, separate testing rooms, etc. to allow candidates to demonstrate their competence. Modification of assessment methods or its administration made to account for the effects of a disability without altering the validity of the assessment.

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 accumulated evidence and theory support specific interpretations of test scores entailed by proposed uses of a test. 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.