

מילון עברי אנגלי למונחים מתחום הפסיכומטריקה

הגדרה S - הגדרה מהסטנדרטים לפיתוח מבחנים N - הגדרה מאתר NCME	מונח אנגלי	מונח עברי
		א
Classification or description of inferred central nervous system status on the basis of neuropsychological assessment. [S]	Neuropsychodiagnosis	אבחון נירוי פסיכולוגי
Formalization or classification of functional mental health status based on psychological assessment. See neuropsychodiagnosis. [S]	Psychodiagnosis	אבחון פסיכולוגי
Limitation of access to the specific content of a test to those who need to know it for test development, test administration, test scoring, or test evaluation. In particular, test items on secure tests are not made available to those who have no legitimate need to see the items; unauthorized copying is forbidden by any test taker or anyone otherwise associated with the test. [S]	Test security	אבטחת מבחן
	Population	אוכלוסייה
	Target population	אוכלוסיית המטרה
The population of test takers represented by test norms. The sample on which the test norms are based must permit accurate estimation of the test score distribution for the reference population. The reference population may be defined in terms of test taker age, grade, or clinical status at time of testing, or other characteristics. [S]	Reference population / Reference group	אוכלוסיית התייחסות / קבוצת התייחסות
	Estimate	אומדן
The manner in which the test questions are presented to the test taker, such as in paper-and-pencil, via a computer terminal, through the internet, or verbally by an examiner. [S]	Test format / Test mode	אופן העברה של מבחן / מתכונת מבחן
	Literacy	אוריינות
Knowledge about testing that supports valid interpretations of test scores for their intended purposes, such as test development practices, test interpretations, threats to valid score interpretations, score reliability and precision, test administration, and use. [S]	Assessment literacy	אוריינות הערכה
The score on a test below which a given percentage of scores for a specified population fall. [S]	Percentile	אחוזון / מאון / מאיון
A program, often legislated, that attributes the responsibility for student learning to teachers, school administrators, and/or students. Test results typically are used to judge accountability, and often consequences are imposed for shortcomings. [N]	Accountability	אחריות
	Invariance	אי-השתנות
	Parameter invariance	אי-השתנות פרמטרים
	Independence	אי-תלות
	Local independence	אי-תלות מקומית

		אינדיקטור – ר' סמן
Strategies that test takers might use while taking the test to improve their performance, such as time management or the elimination of obvious incorrect options on a multiple-choice question before responding to the question. [S]	Test taking strategies	אסטרטגיות היבחנות
	Halo effect	אפקט הילה
	Floor effect	אפקט רצפה
	Ceiling effect	אפקט תקרה
	Response option	אפשרות תגובה
	Testlet	אשכול פריטים
	Testing code of ethics	אתיקה של מפתחי בחינות
ב		
		בחון – ר' משגיח במבחן
Changes to test content, format or administration conditions for particular test takers that do not change the construct being measured but do remove construct -irrelevant contributions to test scores that would otherwise exist for these individuals. Scores from accommodated tests have comparable meaning to those from the original test. [S]	Accommodated testing	בחינה בתנאים מותאמים
		בחינה חוזרת – ר' היבחנות
The aggregate of items from which a test or test scale's items are selected during test development, or the total set of items from which a particular test is selected for a test taker during adaptive testing. [S]	Item bank / Item pool	בנק פריטים / מאגר פריטים
A purpose for testing that results in the acceptance or rejection of applicants for a particular educational or employment opportunity. [S]	Selection	ברירה
ג		
1. Any variable, real or hypothetical, that is an aspect of a concept or construct. 2. In measurement theory, a statistical dimension defined by a factor analysis. See factor analysis. [S]	Factor	גורם
	Group factors	גורמי קבוצה
	Orthogonal factors	גורמים ניצבים
A question or incomplete sentence that poses a problem in a selected-response test item, most often a multiple-choice item. The stem is usually followed by a list of options, which includes distracters and the correct answer, or the keyed response. [N]	Stem	גזע
The question, stimulus, or instructions that direct the efforts of test takers in formulating their responses to a constructed-response exercise. [S]	Item prompt	גריין
ד		
	Random sampling	דגימה מקרית
	Cluster sampling	דגימת אשכולות

A measurement format in which a large set of test items is organized into a number of relatively short item sets, each of which is randomly assigned to a sub-sample of test takers, thereby avoiding the need to administer all items to all test takers. [S]	Matrix sampling	דגימת מטריצה
	Stratified sampling	דגימת שכבות
The process of selecting test items to represent a specified universe of what is intended to be measured. [S]	Domain sampling	דגימת תחום
The characteristic of being relatively proficient in two languages. [S]	Bilingual	דו-לשוני
		דו-סדרתי נקודתי – ל' מקדם מתאם דו-סדרתי
See computer-prepared test interpretation. [S]	Automated narrative report	דוח מילולי ממוחשב / דוח נרטיבי ממוחשב
In classification, diagnosis, or selection, an error in which an individual is assessed or predicted not to meet the criteria for inclusion in a particular group but in truth does (or would) meet these criteria. See sensitivity and specificity. [S]	False negative	דחייה שגויה / החלטה שלילית שגויה
	Precision	דיוק
When test scores are used to assign test takers to specific categories, for example, proficient or not, or a specific diagnosis, classification accuracy is the degree to which the assignment is consistent and accurate, or the degree to which false positive and false negative classifications are avoided. See sensitivity and specificity. [S]	Classification accuracy	דיוק הסיווג
A general term that refers to the impact of measurement error on the outcome of the measurement. See standard error of measurement, error of measurement. [S]	Precision of measurement	דיוק מדידה
	Ranking	דירוג / מדרג
Most commonly, the percentage of scores in a specified distribution that fall below the point at which a given score lies. [S]	Percentile ranking	דירוג אחוזני / דירוג מאוני / דירוג מאיוני
	Response set	דפוס תגובה
ה		
A test characteristic, dictated by the test's time limits, that results in a test taker's score being dependent on the rate at which work is performed as well as the correctness of the responses. The term is not used to describe tests of speed. Speededness is often an undesirable characteristic unless related to the construct being measured. [S]	Speededness	האצה
	Item discrimination	הבחנה של פריט
In testing, the principle that every test-taker should be assessed in an equitable way. Scores from a test that is fair reflect the same construct and have essentially the same meaning for all individuals in the intended test taker population. [S]	Fairness	הוגנות
In terms of educational testing, opportunity to learn addresses whether students have been exposed to the test content through their educational program. [S]	Opportunity to learn	הזדמנות ללמוד
		החלטה שלילית שגויה – ל' דחייה שגויה

In a statistical context, a systematic error in a test score. In discussing test fairness, bias may refer to construct underrepresentation or construct-irrelevant components of test scores that differentially affect the performance of different groups of test takers. See predictive bias, construct underrepresentation, construct irrelevance. [S]	Bias	הטיה
The systematic under- or over-prediction of criterion performance for people belonging to groups differentiated by characteristics not relevant to criterion performance. [S]	Predictive bias	הטיית ניבוי
A test taker's tendency to respond in a particular way or style to items on a test (i.e., acquiescence, social desirability, the tendency to choose 'true' on a true-false test) that yields systematic, construct-irrelevant error in test scores. [S]	Response bias	הטיית תגובה
The test is taken another time; either the same test or, more likely, an alternative form, sometimes with additional training or education between administrations. [S]	Retesting	היבחנות / בחינה חוזרת
Applying validity evidence obtained in one or more situations to other similar situations on the basis of methods such as meta-analysis, or synthetic validation arguments. [S]	Validity generalization	הכללת תוקף
Planned short-term instructional activities for prospective test takers provided prior to the test administration for the primary purpose of improving their test scores. Coaching typically includes practice, instruction on test-taking strategies, and related activities. Activities that approximate the instruction provided by regular school curricula or training programs are not typically referred to as coaching. [S]	Coaching	הכנה
The person(s) or agency responsible for the choice and administration of a test, for the interpretation of test scores produced in a given context, and for any decisions or actions that are based, in part, on test scores. [S]	Test user	המשתמש במבחן
	Rationale	הנמקה
The consistency with which two or more judges rate the work or performance of test takers; sometimes referred to as inter-rater reliability. [S]	Inter-rater agreement	הסכמה בין מעריכים / הסכמה בין שופטים
The agreement of a person, or that person's legal representative, for some procedure to be performed on or by the individual, such as taking a test or completing a questionnaire. The agreement, which is usually written, is made after the nature, possible effects, and use of the procedure has been explained. [S]	Informed consent	הסכמה מדעת
A process, often national in scope, by which individuals who have been certified have demonstrated some level of knowledge and skill in an occupation. See licensing, credentialing. [S]	Certification	הסמכה
Granting to a person, by some authority, a credential, such as a certificate, license, or diploma, that signifies an acceptable level of performance in some domain of knowledge or activity. [S]	Credentialing	הסמכה
	Inference	הסקה
	Probability	הסתברות
	Test administration	העברת מבחן / מתן מבחן

Any systematic method of obtaining information from tests and other sources, used to draw inferences about characteristics of people, objects, or programs. [S]	Assessment	הערכה
The process of gathering information to make a judgment about the quality or worth of some program or performance. The term also is used to refer to the judgment itself, as in "My evaluation of his work is . . ." [N]	Evaluation	הערכה
An assessment containing items that are judged to be measuring the ability to apply and use knowledge in real-world contexts. [N]	Authentic assessment	הערכה אותנטית
	Scholastic assessment	הערכה אקדמית
	Behavioral assessment	הערכה התנהגותית
	External evaluation	הערכה חיצונית
Assessments or tests used to evaluate the performance of students who are unable to participate in general assessments even with accommodations. Alternate assessments address alternate standards and provide a mechanism for students with the most significant cognitive disabilities, and for other students with disabilities who may need alternate ways to access assessments, to be included in an educational accountability system. [S]	Alternate assessment / Alternative assessment	הערכה חלופית
	Forensic evaluation	הערכה לצרכי משפט
The assessment of a student's knowledge and skills at a particular point in time. It is typically carried out at the completion of a program of learning, such as the end of an instructional year in school. [S]	Summative assessment	הערכה מסכמת
Assessments administered periodically throughout the school year, at specified times during a curriculum sequence, to evaluate students' knowledge and skills relative to an explicit set of longer-term learning goals. [S]	Benchmark assessment	הערכה מעוגנת נקודות קבע / הערכה מעוגנת ספי הישג
Assessments of an individual's standing with respect to systematically described content and performance standards. [S]	Standards-based assessment	הערכה מעוגנת תקנים
A process used by teachers and students during instruction that provides feedback to adjust ongoing teaching and learning to improve students' achievement of intended instructional outcomes. [S]	Formative assessment	הערכה מעצבת
A specialized type of psychological assessment of normal or pathological processes affecting the central nervous system and the resulting psychological and behavioral functions or dysfunctions. [S]	Neuropsychological assessment	הערכה נירופסיכולוגית
An assessment instrument for which the respondent answers questions or makes ratings of his/her own behavior or performance, as opposed to such responses being made by an observer of that individual. (See also inventory.) [N]	Self-evaluation	הערכה עצמית
	Internal evaluation	הערכה פנימית
A comprehensive examination of psychological functioning that involves collecting, evaluating, and integrating test results and collateral information, and reporting information about an individual. Psychological assessments are undertaken to answer specific questions about a client's psychological functioning during a particular time interval or to predict a client's psychological functioning in the future. [S]	Psychological assessment	הערכה פסיכולוגית

The process of systematically gathering test scores and related data in order to make judgments about an individual's ability to perform various mental activities involved in the processing, acquisition, retention, conceptualization, and organization of sensory, perceptual, verbal, spatial, and psychomotor information. [S]	Cognitive assessment	הערכה קוגניטיבית
	Clinical assessment	הערכה קלינית
Assessments administered during instruction to evaluate students' knowledge and skills relative to a specific set of academic goals in order to inform policymaker or educator decisions at the classroom, school or district level. [S]	Interim assessments	הערכת ביניים
Product- and behavior-based measurements based on settings designed to emulate real-life contexts or conditions in which specific knowledge or skills are actually applied. [S]	Performance assessments	הערכת ביצוע
The measurement of an incumbent's performance of a job. This may include a job sample test, an assessment of job knowledge, or ratings of the incumbent's actual performance on the job. [S]	Job performance measurement	הערכת ביצוע תפקיד
A specialized type of psychological assessment designed to generate hypotheses and inferences about interests, work needs and values, career development, vocational maturity, and indecision. [S]	Vocational assessment	הערכת התאמה לעיסוק
	Peer evaluation	הערכת עמיתים
An evaluation of the efficacy of an intervention. [S]	Outcome evaluation	הערכת תוצאה
1. The collection and synthesis of systematic evidence about the use, operation, and effects of a program. 2. The set of procedures used to make judgments about a program's design, its implementation, and its outcomes. [S]	Program evaluation	הערכת תכנית
The process of collecting data and making an appraisal of a person's ability to focus on the relevant stimuli in a situation. The assessment may be directed at mechanisms involved in arousal, sustained attention, selective attention and vigilance, or limitation in the capacity to attend to incoming information. [S]	Attention assessment	הערכת תפקודי קשב
	Randomization	הקצאה מקרית
The consequences of testing refers to the outcome, intended and unintended, of using tests in particular ways in certain contexts and with certain populations. [S]	Consequences	השלכות
A test designed to determine which course, in a sequence of courses, would be optimal for a student to enroll in to begin study. These tests often are used by colleges to determine which of several mathematics, chemistry, or foreign language courses is the best starting place for a student who has taken courses in these areas in high school. [N]	Placement	השמה
	Response omission	השמטת תגובה
Influence of item position, other items administered, time limits, administration conditions, etc. on item difficulty and other statistical item characteristics. Item context effects are especially important to control when items are used as anchor items in equating. [S]	Item context effect	השפעת הקשר הפריט
	Covariance	השתנות משותפת

Changes to test content, format or administration conditions for particular test takers that do not change the construct being measured but do remove construct-irrelevant contributions to test scores that would otherwise exist for these individuals. Scores from accommodated tests have comparable meaning to those from the original test. [S]	Accommodations	התאמות / תנאים מותאמים
Changes made in the content, format, and/or administration procedure of a test to increase the accessibility of the test for test takers who are unable to take the original test under standard test conditions. Test modifications change the construct being measured by the test and hence score interpretations. See modifications, test modifications. [S]	Test modifications	התאמות במבנה מבחן / שינויים במבנה מבחן
Adaptation, as used in the Standards, is any change in test content, format, or administration conditions that is made to increase the test accessibility for individuals who otherwise would face construct-irrelevant barriers on the original test. An adaptation may or may not change the meaning of the construct being measured or alter score interpretations. An adaptation that changes score meaning is referred to as a modification; an adaptation that does not is referred to as an accommodation. See accommodations and modifications. [S]	Test adaptation	התאמת מבחן (לאוכלוסייה ספציפית)
	Model fitting	התאמת מודל
In test linking, the term moderation, used without a modifier, usually signifies statistical moderation, which is the process of relating scores on one test to scores on another, so that scores have the same relative meaning for a group of test takers. [S]	Moderation	התאמת ציונים
In test linking, projection is a method of linking in which scores on one test are used to predict scores on another test for a group of test takers, often using regression methodology. [S]	Projection	התאמת ציונים
	Faking	התחזות
Any of a number of activities in which a prospective test taker might participate, primarily for the purpose of optimizing their score on an upcoming test. The nature of the activities, the circumstances under which they are presented, and the purpose of the test must all be considered in deciding whether specific test preparation is either appropriate and beneficial or inappropriate and unethical or merely unhelpful. (See also coaching.) [N]	Test preparation	התכוננות למבחן
	Convergence	התכנסות
	Distribution	התפלגות
A graphic display of scores for a large group that has the shape of a bell: many persons have scores in the middle and a much smaller number have very high or very low scores. Many physical and psychological characteristics demonstrate a normal, or bell-shaped, distribution when shown graphically. (It is often called the normal curve.) [N]	Normal distribution	התפלגות נורמלית
ז		
	Latency of responses	זמן חביון
	Reaction time	זמן תגובה

The extent to which test scores are influenced by extraneous factors that are irrelevant to the construct that the test is intended to measure and lead to errors in interpretation, distorting the meaning of these scores. [S]	Construct irrelevance	זרות למבנה
ח		
	Unidimensionality	חד-ממדיות
The amount of skill in test taking possessed by an individual. The skill relates to such things as time management, how to guess among options when the test taker has partial or little knowledge about the ideas in a test item, how to provide constructed responses that might be most appealing to scorers, and how to identify unintended cues in items prepared by less-experienced test developers. [N]	Test wiseness	חוכמת היבחנות
	Lawfulness	חוקיות
		חזאי – ר' מבא
		חיזוי – ר' נבו
The score above which and below which exactly half of the scores in a certain group are located when the scores are placed in order from high to low. It also is known as the middle score or 50th percentile. [N]	Median	חציון
	Test anxiety	חרדת בחינות
ט		
	Common range	טווח משותף
	Ordered range	טווח סדר
	Range of scores / Score range	טווח ציונים
An explicit justification of the degree to which accumulated evidence and theory support the proposed interpretation(s) of test scores for their intended uses. [S]	Validity argument	טיעון תומך תוקף
The standard deviation of an individual's observed scores from repeated administrations of a test (or parallel forms of a test) under identical conditions. Because such data cannot generally be collected, the standard error of measurement is usually estimated from group data. See error of measurement. [S]	Standard error of measurement	טעות התקן של המדידה
The difference between an observed score and the corresponding true score or proficiency. See standard error of measurement and true score. [S]	Error of measurement	טעות מדידה
An unsystematic error; a quantity (often observed indirectly) that appears to have no relationship to any other variable. [S]	Random error	טעות מקרית
A consistent score component (often observed indirectly), not related to the test performance. See bias. [S]	Systematic error	טעות שיטתית
The standard deviation of measurement errors that affect the scores of test takers at a specified test score level. [S]	Conditional standard error of measurement	טעות תקן מותנית של מדידה
	Factor loading	טעינות גורם
י		
		ייחודיות – ר' ספציפיות
	Selection ratio	יחס ברירה
	Uniqueness	ייחודיות

The extent to which a test fails to capture important aspects of the construct that the test is intended to measure. In this situation, the meaning of test scores is narrower than the proposed interpretation implies. [S]	Construct underrepresentation	ייצוג חסר של מבנה / ייצוג חסר של תכונה
The extent to which scores on a test are essentially invariant over time. Stability is an aspect of reliability and is assessed by correlating the test scores of a group of individuals with scores on the same test, or an equated test, taken by the same group at a later time. [S]	Stability	יציבות
כ		
1. In linking test scores, the process of relating scores on one test to scores on another, so that scores have the same relative meaning for a group of test takers. 2. In item response theory, the process of estimating the parameters of the item response function. 3. A term used to describe the process of promoting agreement in scoring among individual raters of a set of performance outcomes. [S]	Calibration	כיול
Relating scores on two or more alternate forms. The equated scores are typically reported on a common score scale. [S]	Equating	כיול
	Vertical equating	כיול אנכי
	Selection rule	כלל ברירה (סטטיסטי, סמנטי)
	Construction rule	כלל יצירה
	Faculties	כשרים
מ		
		מאגר פריטים – ל' בנק פריטים
		מאון – ל' אחוזון
		מאון – ל' אחוזון
A statistic that describes how much the scores in a particular group vary; it is a measure of variability. Some statistical techniques used in testing depend on being able to partition the variance and attribute the parts to various test, test administration, or test taker characteristics. Statistically, it is also the same as the standard deviation squared. [N]	Variability	משתנות
An evaluative device or procedure in which a sample of a test taker's behavior in a specified domain is obtained and subsequently evaluated and scored using a standardized process. [S]	Test	מבחן
A sequential form of individual testing in which successive items, or sets of items, in the test are chosen based primarily on their psychometric properties and content, in relation to the test taker's responses to previous items. [S]	Adaptive test	מבחן אדפטיבי / מבחן מסתגל
An adaptive test administered by computer. See adaptive testing. [S]	Computerized adaptive test	מבחן אדפטיבי ממוחשב / מבחן מסתגל ממוחשב
	Culture-fair test	מבחן הוגן תרבותית
	Homogeneous test	מבחן הומוגני

A test to evaluate the extent of knowledge or skill attained by a test taker in a content domain in which the test taker had received instruction. [S]	Achievement test	מבחן הישגים
A test administered to a sample of test takers to try out some aspects of the test or test items, such as instructions, time limits, item response formats, or item response options. See field test. [S]	Pilot test	מבחן חלוץ / מבחן ניסיוני
The use of standardized tests to evaluate the current performance of a person in some defined domain of cognitive, psychomotor, or physical functioning. [S]	Ability test	מבחן יכולת
A test designed and used to predict how well someone might perform in a certain ability area in the future. Examples include scholastic, musical, clerical, verbal, and mechanical aptitude. [N]	Aptitude test	מבחן כישורים
A test that is used to make broad categorizations of test takers as a first step in selection decisions or diagnostic processes. [S]	Screening test	מבחן לסינון ראשוני
A test of the ability of an individual to perform the tasks of which the job is comprised. [S]	Job sample test	מבחן מדמה תפקיד
	Speeded test	מבחן מואץ
A test administered to a test taker who is allotted a strictly prescribed amount of time to respond to the test. [S]	Timed test	מבחן מוגבל בזמן
A change in test content, format and/or administration conditions that is made to increase accessibility for some individuals but which affects the construct measured and, consequently, result in scores that differ in meaning from scores from the unmodified assessment. Because modified assessments are measuring a different construct from that measured by the standardized assessment, it is important to interpret the assessment scores as resulting from a new test and to gather whatever empirical evidence is necessary to determine the validity of the interpretations for intended uses of the scores. [S]	Modified test	מבחן מוגש
A test used to provide results that have only minor or indirect consequences for test takers, programs, or institutions involved in the testing. [S]	Low-stakes test	מבחן מועט-סיכון
A test administered by a computer. Questions appear on a computer-produced display, and the test taker responds by using a keyboard, "mouse," or other similar response device. [S]	Computer-administered test / Computer-based test	מבחן ממוחשב
		מבחן מסתגל – ר' מבחן אדפטיבי
		מבחן מסתגל ממוחשב – ר' מבחן אדפטיבי ממוחשב
A test that allows its users to make score interpretations in relation to a performance level, as distinguished from those interpretations that are made in relation to the performance of others. Examples of criterion-referenced interpretations include comparison to cut scores, interpretations based on expectancy tables, and domain-referenced score interpretations [S]	Criterion-referenced test	מבחן מעוגן קריטריון
A psychological or educational test designed to measure an individual's level of cognitive functioning in accord with some recognized theory of intelligence. [S]	Intelligence test	מבחן משכל

		מבחן ניסיוני – ר' מבחן חלוץ
	Culture-free test	מבחן נקי מהשפעה תרבותית
A set of anchor items. [S]	Anchor test	מבחן עוגן
A test used to provide results that have important, direct consequences for test takers, programs, or institutions involved in the testing. [S]	High-stakes test	מבחן עתיר-סיכון
tests that are administered to groups of test takers, usually in a group setting, typically with standardized administration procedures and supervised by a proctor or test administrator. [S]	Group test	מבחן קבוצתי
A test administration used to check the adequacy of testing procedures, and the statistical characteristics of new test items or new test forms. A field test is generally more extensive than a pilot test. See pilot test. [S]	Field test	מבחן שדה
1. A criterion-referenced test designed to indicate the extent to which the test taker has mastered some domain of knowledge or skill. Mastery is generally indicated by attaining a passing score or cut score. 2. In some technical use, a test designed to indicate whether a test taker has or has not attained a prescribed level of mastery of a domain. See cut score, computer-based mastery test. [S]	Mastery test	מבחן שליטה
Tests that are administered because of a mandate from an external authority. [S]	Mandated tests	מבחני חובה
	Educational testing	מבחנים חינוכיים
Any procedure that involves the use of tests or inventories to assess particular psychological characteristics of an individual. [S]	Psychological tests	מבחנים פסיכולוגיים
	Large scale assessments	מבחנים רחבי היקף
1. The set of factors obtained in a factor analysis. 2. Technically, the correlation of each factor with each of the original variables from which the factors are derived. [S]	Factorial structure	מבנה גורמי
In test analysis, the factorial structure of item responses or subscales of a test. See factorial structure. [S]	Internal structure	מבנה פנימי
	Item format	מבנה פריט
A selection of a specified number of entities called sampling units (test takers, items, etc.) from a larger specified set of possible entities, called the population. A random sample is a selection according to a random process, with the selection of each entity in no way dependent on the selection of other entities. A stratified random sample is a set of random samples, each of a specified size, from several different sets, which are viewed as strata of the population. [S]	Sample	מדגם
See sample. [S]	Random sample	מדגם מקרי
See sample. [S]	Stratified sample	מדגם שכבות
	Standardization sample	מדגם תקנון
	Goodness of fit index	מדד טיב התאמה
	Proximity measure	מדד קרבה
	Central tendency, measures of	מדדי נטייה מרכזית
	Association, measures of	מדדי קשר

The process of assigning a number to a person, or a person's trait, according to specified rules. Often the rules involve using a test and counting the number of items each person answered correctly. That number represents how much of the trait the person has, and it can be compared with other information to obtain further meaning about their performance. [N]	Measurement	מידה
The principles, plan, or procedures established by an agency, institution, organization, or government, generally with the intent of reaching a long-term goal. [S]	Policy	מדיניות
		מדרג – ר' דירוג
A publication prepared by test authors and publishers to provide technical and psychometric information on a test. [S]	Technical manual	מדריך טכני
A publication prepared by the test authors and publishers to provide information on a test's purpose, appropriate uses, proper administration, scoring procedures, normative data, interpretation of results, and case studies. See test manual. [S]	User's guide	מדריך למשתמש
A publication prepared by test developers and publishers to provide information on test administration, scoring, and interpretation and to provide technical data on test characteristics. See user's guide. [S]	Test manual	מדריך למשתמש במבחן
The degree to which test scores for a group of test takers are consistent over repeated applications of a measurement procedure and hence are inferred to be dependable, and repeatable for an individual test taker; the degree to which scores are free of errors of measurement for a given group. See generalizability theory. [S]	Reliability	מהימנות
The degree of agreement among repetitions of a single rater scoring students' responses. Inconsistencies in the scoring process resulting from influences that are internal to the rater rather than true differences in student performance result in low intra-rater reliability. [S]	Intra-rater reliability	מהימנות בין מעריכים / מהימנות בין שופטים
A reliability coefficient obtained by administering the same test a second time to the same group after a time interval and correlating the two sets of scores; typically used as a measure of stability of the test scores. [S]	Test-retest reliability	מהימנות מבחן חוזר
	Hierarchical model	מודל היררכי
		מהימנות מותאם – ר' מקדם תוקף
	Value added model	מודל הערך המוסף
	Conjunctive model	מודל מאחד
	Compensatory model	מודל מפצה
	Disjunctive model	מודל מפצה
A statistical model used to estimate the progress of test takers over time and different assessment occasions. [S]	Growth model	מודל צמיחה
	Structural equation models	מודלים של משוואות מבניות
	Conditional	מותנה
See scoring rubric. [S]	Rubric / Scoring rubric	מחון
Research that addresses ease of use, of how quickly someone can understand how to use something and how easily they can use it. [S]	Usability research	מחקר שימושיות

An assessment tool that requires test takers to perform—develop a product or demonstrate a process—so that the observer can assign a score or value to that performance. A science project, an essay, a persuasive speech, a mathematics problem solution, and a woodworking project are examples. (See also authentic assessment.) [N]	Performance task	מטלת ביצוע
	Multitrait-multimethod matrix	מטריצה מרובת תכונות ושיטות
	Skill	מיומנות
		מיון – ר' סיווג
	Universe of items	מכלול של פריטים
The average score obtained by some identified group. All scores are added and the sum is divided by the number of scores. (See also mode and median.) [N]	Mean	ממוצע
	Grade point average	ממוצע ציונים
	Predictor	מנבא / חזאי
Historically, a score obtained by dividing a person's mental age score, obtained by administering an intelligence test, by the person's chronological age, both expressed in terms of years and months. The resulting fraction is multiplied by 100 to obtain the IQ score. [N]	I.Q.	מנת משכל
The incorrect options that are listed with the keyed response in a multiple-choice or other selected-response test item. Sometimes called foils. [N]	Distractor	מסיח
Documents such as test manuals, technical manuals, user's guides, specimen sets, and directions for test administrators and scorers that provide information for evaluating the appropriateness and technical adequacy of a test for its intended purpose. [S]	Test documents	מסמכים נלווים למבחן
The chronological age in a defined population for which a given score is the median (middle) score. Thus, if children 10 years and 6 months of age have a median score of 17 on a test, the score 17 is said to have an age equivalent of 10-6 for that population. See grade equivalent. [S]	Age equivalent	מעוגן גיל / תואם גיל
The school grade level for a given population for which a given score is the median score in that population. See age equivalent. [S]	Grade equivalent	מעוגן דרגת כיתה / תואם דרגת כיתה
	Slipping	מעידה
	Study design	מערך מחקר
The person(s) or agency responsible for the construction of a test and for the documentation regarding its technical quality for an intended purpose. [S]	Test developer	מפתח מבחנים
A detailed description for a test, often called a test blueprint, that specifies the number or proportion of items in the test that assess each content and process/skill area; the format of items, responses, and scoring rubrics and procedures; and the desired psychometric properties of the items and test such as the distribution of item difficulty and discrimination indices. [S]	Test specifications	מפרט מבחן

An index of reliability/precision based on generalizability theory (G theory). A generalizability coefficient is the ratio of universe score variance to observed score variance, where the universe score variance is equal to the observed score variance plus the total error variance. See generalizability theory. [S]	Generalizability coefficient	מקדם ההכללה
	Coefficient of alienation (COA)	מקדם הזרות
An internal consistency reliability coefficient based on the number of parts into which the test is partitioned (e.g., items, subtests, or raters), the interrelationships of the parts, and the total test score variance. Also called Cronbach's alpha and, for dichotomous items, KR 20. [S]	Coefficient alpha	מקדם המהימנות אלפא
A unit-free indicator that reflects the degree to which scores are free of measurement error. In classical test theory, the term represents the ratio of true score variance to observed score variance for a particular test taker population. The conditions under which the coefficient is estimated may involve variation in test forms, measurement occasions, raters, scorers, or clinicians, and may entail multiple test taker products or performances. These and other variations in conditions give rise to qualifying adjectives, such as alternate-form reliability, internal consistency reliability, test-retest reliability, etc. See generalizability theory. [S]	Reliability coefficient	מקדם מהימנות
An internal consistency coefficient obtained by using half the items on the test to yield one score and the other half of the items to yield a second, independent score. The correlation between the scores on these two half-tests, adjusted via the Spearman-Brown formula, provides an estimate of the alternate-form reliability of the total test. [S]	Split-halves reliability coefficient	מקדם מהימנות של מבחן חצוי
	Biserial/Point-biserial correlation coefficient	מקדם מתאם דו-סדרתי / דו-סדרתי נקודתי
An index of the reliability of test scores derived from the statistical interrelationships of responses among item responses or scores on separate parts of a test. [S]	Internal consistency coefficient	מקדם עקיבות פנימית
A validity or reliability coefficient—most often, a product-moment correlation—that has been adjusted to offset the effects of differences in score variability, criterion variability, or the unreliability of test and/or criterion. See restriction of range or variability. [S]	Adjusted validity/reliability coefficient	מקדם תוקף/מהימנות מותאם
	Universe of admissible observations	מרחב התצפיות הקבילות
	Assessment center	מרכז הערכה
Persons responsible during the test administration for monitoring the testing process and ensuring that the testing protocol is followed. [S]	Test proctor	משגיח במבחן / בוחן
	Mapping sentence	משפט מיפוי
	Weight	משקל
	Variable	משתנה
	Endogenous variable	משתנה אנדוגני
	Exogenous variable	משתנה אקסוגני
	Independent variable	משתנה בלתי תלוי
	Latent variable	משתנה חבוי
	Suppressive variable	משתנה מדכא /

		משתנה מעכב
	Confounding variable	משתנה ממסך
		משתנה מעכב – ר' משתנה מדכא
	Mediating variable	משתנה מתווך
A variable that affects the direction or strength of the relationship between two other variables. [S]	Moderator variable	משתנה מתערב
	Dependent variable	משתנה תלוי
A statistic used to show how the scores from one measure relate to scores on a second measure for the same group of individuals. A high value (approaching +1.00) is a strong direct relationship, a low negative value (approaching -1.00) is a strong inverse relationship, and values near 0.00 indicate little, if any, relationship. [N]	Correlation	מתאם
	Multicollinearity	מתאם גבוה בין-חזאים
	Biserial correlation	מתאם דו-סדרתי
	Point-biserial correlation	מתאם דו-סדרתי נקודתי
	Partial correlation	מתאם חלקי
		מתכונת התגובה – ר' סוג התגובה
		מתכונת מבחן – ר' אופן העברה של מבחן
		מתן מבחן - ר' העברת מבחן
נ		
As used in educational assessment, accessibility refers to the degree to which the items or tasks on a test enable as many test takers as possible to demonstrate their standing on the target construct without being impeded by characteristics of the item that are irrelevant to the construct being measured. [S]	Accessibility	נגישות
Two or more versions of a test that are considered interchangeable, in that they measure the same constructs in the same ways, are built to the same content and statistical specifications and are administered under the same conditions using the same directions. Sometimes referred to as equivalent forms or parallel forms. [S]	Alternate forms	נוסחים חלופיים
Alternate forms whose scores have been related through statistical equating. Following an appropriate equating, scale scores on equated forms can be used interchangeably. [S]	Equated forms	נוסחים מכילים
See alternate forms. [S]	Parallel forms	נוסחים מקבילים
See alternate forms. [S]	Equivalent forms	נוסחים שקולים
The formula by which the raw score on a test is obtained. The simplest scoring formula is "raw score equals number correct." Other formulas differentially weight item responses. For example, in an attempt to correct for guessing or nonresponse, zero weights may be assigned to nonresponses and negative weights to incorrect responses. [S]	Scoring formula	נוסחת נקידה

Statistics or tabular data that summarize the distribution or frequency of test scores for one or more specified groups, such as test takers of various ages or grades. Norms are usually designed to represent some larger population, such as test takers throughout the country. The group of test takers represented by the norms is referred to as the reference population. [S]	Norms	נורמות
	Age norms	נורמות גיל
Norms by which test scores are referred to a specific, limited reference population of particular interest to the test user (e.g., locale, organization, or institution); local norms are not intended as representative of populations beyond that setting. [S]	Local norms	נורמות מקומיות
Descriptive statistics (including percentile ranks) for a sample of test takers that does not represent a well-defined reference population, for example, all persons tested during a certain period of time, or a set of self-selected test takers. See local norms; norms. [S]	User norms	נורמות מקומיות
	Prediction	ניבוי / חיזוי / תחזית
	Guessing	ניחוש
	Analysis	ניתוח
A statistical method of research in which the results from several independent, comparable studies are combined to determine the size of an overall effect or the degree of relationship between two variables. [S]	Meta-analysis	ניתוח-על
Any of several statistical methods of describing the interrelationships of a set of variables by statistically deriving new variables, called factors, that are fewer in number than the original set of variables. [S]	Factor analysis	ניתוח גורמים
	Exploratory analysis	ניתוח חקרני / ניתוח מגשש
	Confirmatory analysis	ניתוח מאשש
	Tendency analysis	ניתוח מגמה / ניתוח נטייה
		ניתוח מגשש – ר' ניתוח חקרני
		ניתוח נטייה – ר' ניתוח מגמה
	Path analysis	ניתוח נתיבים
		ניתוח עיסוק – ר' ניתוח תפקיד
A procedure used by test developers to examine the quality of an item prior to its selection for use on a test, or to determine how the item might be revised before its subsequent selection. Often statistical properties such as difficulty, discrimination, and DIF are evaluated in the process. [N]	Item analysis	ניתוח פריטים
	Utility analysis	ניתוח תועלת
A general term referring to the investigation of positions or job classes to obtain information about job duties and tasks, responsibilities, necessary worker characteristics (e.g. knowledge, skills, and abilities), working conditions, and/or other aspects of the work. [S]	Job analysis / Practice Analysis	ניתוח תפקיד / ניתוח עיסוק
	Multiple cutoffs	נקודות חתך

Assessments administered periodically throughout the school year, at specified times during a curriculum sequence, to evaluate students' knowledge and skills relative to an explicit set of longer-term learning goals. [S]	Benchmarks	נקודות קבע / ספי הישג
	Scoring	נקידה / ציון
A method of scoring in which each critical dimension of performance is judged and scored separately, and the resultant values are combined for an overall score. In some instances, scores on the separate dimensions may also be used in interpreting performance. See holistic scoring. [S]	Analytic scoring	נקידה אנליטית / נקידה מבוססת ממדים
A method of obtaining a score on a test, or a test item, based on a judgment of overall performance using specified criteria. See analytic scoring. [S]	Holistic scoring	נקידה הוליסטית
		נקידה מבוססת ממדים – ל' נקידה אנליטית
A procedure by which constructed response items are scored by computer using an algorithmically-based approach. Automated scoring procedures usually require responses that are based on a computer-administered test. [S]	Automated scoring / Computerized scoring	נקידה ממוחשבת
	Formula scoring	נקידה על-פי נוסחה
	Likelihood	נראות
O		
	Tolerance	סבילות
The mechanism that a test taker uses to respond to the test question by either selecting from a list of options (multiple-choice questions) or providing a written response (fill-in, verbal or written response to an open or constructed response question). [S]	Response format	סוג התגובה / מתכונת התגובה
	Item type	סוג פריט
A set of tests usually administered as a unit. The scores on the tests usually are scaled so that they can readily be compared or used in combination for decision making. [S]	Battery	סוללת מבחנים
1. The system of numbers, and their units, by which a value is reported on some dimension of measurement. 2. In testing, scale sometimes refers to the set of items or subtests used in the measurement and is distinguished from a test in the type of characteristic being measured. One speaks of a test of verbal ability, but a scale of extroversion-introversion. [S]	Scale	סולם
	Forced-choice scale	סולם בחירה כפויה
	Likert scale	סולם ליקרט
	Ratio scale	סולם מנה
	Ordinal scale	סולם סדר
A subset of items, common to two or more test forms, that are administered for purposes of score equating. (See also equating.) [N]	Anchor scale	סולם עיגון
	Interval scale	סולם רווח
	Nominal scale	סולם שמי
	Scales of measurement	סולמות מדידה
	Item parameter drift	סחף פרמטרים של פריט
	Deviance	סטייה

A statistic that describes how much the scores in a particular group vary; it is a measure of variability. Conceptually, the number indicates the average amount by which the scores in a group differ from their mean score. It is also the square root of the variance. [N]	Standard deviation	סטיית תקן
	Classification	סיווג / מיון
The process of creating a scale or a scale score. Scaling is intended to enhance test score interpretation by placing scores from different tests or test forms onto a common scale or by producing scale scores designed to support criterion-referenced or norm-referenced score interpretations. [S]	Scaling	סילום
In test linking, vertical scaling is the process of relating scores on tests that measure the same construct but differ in difficulty. Vertical scaling is typically used with achievement and ability tests with content or difficulty that spans a variety of grade or age levels. [S]	Vertical scaling	סילום אנכי
	Contiguity	סמיכות
	Indicator	סמן / אינדיקטור
	Threshold	סף
		ספי הישג - ר' נקודות קבע
In classification of disorders, the proportion of cases for which a diagnosis of disorder is rejected when rejection is warranted. [S]	Specificity	ספציפיות / ייחודיות
ע		
Evidence based on some form of data, as opposed to that based on logic or theory. [S]	Empirical evidence	עדות אמפירית
Evidence (usually related to reliability or validity) collected for a specific set of test takers in a single institution or at a specific location. [S]	Local evidence	עדות מקומית
	Synthetic validity evidence	עדות תוקף משולבת
	Content universe	עולם תוכן
	Attitude	עמדה
	Guessing penalty	ענישה על ניחוש
A mathematical function relating the probability of a certain item response, usually a correct response, to the level of the attribute measured by the item. Also called item response curve, or item response function, or icc. [S]	Item characteristic curve	עקומת אופיין הפריט
	Bell curve	עקומת פעמון
	Plausible values	ערכים סבירים
פ		
	Function	פונקציה / תפקוד
	Logistic function	פונקציה לוגיסטית
	Information function	פונקציית אינפורמציה / פונקציית מידע
A mathematical function relating each level of an ability or latent trait, as defined under item response theory (IRT), to the reciprocal of the corresponding conditional measurement error variance. [S]	Test information function	פונקציית האינפורמציה של המבחן

	Item response function	פונקציית התגובה לפריט / אופיין פריט
	Step function	פונקציית מדרגה
A score interpretation based on a comparison of a test taker's performance to the performance of other people in a specified reference population. See criterion-referenced score interpretation. [S]	Norm-referenced score interpretation	פירוש ציון מעוגן נורמות
The process through which a test is planned, constructed, evaluated, and modified, including consideration of content, format, administration, scoring, item properties, scaling, and technical quality for its intended purpose. [S]	Test development / Test construction	פיתוח מבחנים
The process of developing detailed specifications for what a test is to measure and the content, cognitive level, format, and types of test items to be used. [S]	Test design	פיתוח מפרט מבחן
Literally, the term refers to psychological measurement. Generally, it refers to the field in psychology and education that is devoted to testing, measurement, assessment, and related activities. [N]	Psychometrics	פסיכומטריקה
A record of the responses given by a test taker to a particular test. [S]	Response protocol	פרוטוקול תגובה
A statement, question, exercise, or task on a test for which the test taker is to select or construct a response, or perform a task. See item prompt. [S]	Item	פריט
	Multiple choice item	פריט ברירה
	Matching item	פריט התאמה
	Essay item	פריט חיבור / פריט כתיבת חיבור
	Cloze item	פריט מילוי / פריט קלז
An exercise or task for which test takers must create their own responses or products rather than choose a response from an enumerated set. Short- answer items require a few words or a number as an answer, whereas extended-response items require at least a few sentences and may include diagrams, mathematical proofs, essays, etc. [S]	Constructed response item	פריט פתוח / פריט מסוג הפקת תגובה
A test item that requires the test taker to furnish a word, phrase, sentence, or numerical response. These are constructed-response items that require only a brief response rather than several paragraphs or pages (an extended response). [N]	Short answer item	פריט תשובה קצרה
Items that are administered with each of two or more alternate forms of a test for the purpose of equating the scores obtained on these alternate forms. [S]	Anchor items	פריטי עוגן
	Parameter	פרמטר
In item response theory (IRT), a theoretical value indicating the level of a test taker on the ability or trait measured by the test; analogous to the concept of true score in classical test theory. [S]	Ability parameter	פרמטר היכולת
	Score interpretation	פרשנות הציונים

A summary, or a set of illustrations, of the intended meaning of test scores, based on the construct(s) or concept(s) the test is designed to measure. [S]	Proposed interpretation	פרשנות מוצעת (לציון)
The meaning of a test score for an individual or an average score for a defined group, indicating an individual's or group's level of performance in some defined criteria on domain. [S]	Criterion-referenced score interpretation	פרשנות ציון מעוגנת קריטריון
צ		
	Item bundle	צביר פריטים
Any specific number resulting from the assessment of an individual; a generic term applied for convenience to such diverse measures as a, raw score, scale score, estimate of a latent variable, a production count, an absence record, a course grade, a rating, and so forth. [S]	Score / Grade	ציון
In classical test theory, the average of the scores that would be earned by an individual on an unlimited number of perfectly parallel forms of the same test. In item response theory, the error-free value of test taker proficiency, usually symbolized by θ . [S]	True score	ציון אמתי
The score on a test that is often calculated by counting the number of correct answers, but more generally a sum or other combination of item scores. In item response theory, the estimate of test taker proficiency is analogous to a raw score. [S]	Raw score	ציון גלם
	Difference score	ציון הפרש
A specified point on a score scale, such that scores at or above that point are reported, interpreted or acted upon differently from scores below that point. [S]	Cut score	ציון חתך
In testing, the difference between two scores obtained by a test taker on the same test or two equated tests taken on different occasions, often before and after some treatment. [S]	Gain score	ציון מוסף
A score found by transforming raw scores. Scale scores may aid interpretation by indicating how a given score compares to those of other test takers, by enhancing the comparability of scores obtained using different forms of a test, to avoid confusion with other scores, or in other ways [S]	Scale score	ציון מסולם
A score that combines several scores according to a specified formula. [S]	Composite score	ציון מצרף
A method of scoring a test in which the number of points awarded for a correct (or diagnostically relevant) response is not the same for all items in the test. In some cases, the scoring formula awards more points for one response to an item than for another. [S]	Weighted scoring	ציון משוקלל
	Sum score / Total score	ציון סכום / ציון כולל
	Polytomous score	ציון פוליטומי / ציון רב-ערכי

A type of scale score such that the distribution of these scores for a specified population has convenient, known values for the mean and standard deviation. The term is sometimes used to signify a mean of 0.0 and a standard deviation of 1.0. See scale score. [S]	Standard score	ציון תקן
A transformed test score in which a numerical transformation has been chosen so that the transformed score distribution closely approximates a normal distribution, for some specific population. [S]	Normalized standard score	ציון תקן מנורמל
	Normative scores	ציוני נורמה
		ציינון – ר' נקידה
ק		
	Contrasting groups	קבוצות מנוגדות
		קבוצת התייחסות – ר' אוכלוסיית התייחסות
	Focal group	קבוצת מטרה / קבוצת מיקוד
The process of setting cut scores using a structured procedure. Often procedures are judgmentally based and incorporate consideration of performance levels and performance level descriptors. When feasible, standard setting methods are informed by sound empirical data concerning the relation of test performance to the relevant criteria. [S]	Standard setting	קביעת תקנים
In classification, diagnosis, or selection, an error in which an individual is assessed or predicted to meet the criteria for inclusion in a particular group but in truth does not (or would not) meet these criteria. See sensitivity and specificity. [S]	False positive	קבלה שגויה / החלטה חיובית שגויה
	Item difficulty	קושי פריט
	Category	קטגוריה
Reduction in the observed score variance of a test taker sample, compared to the variance of the entire test taker population, as a consequence of constraints on the process of sampling test takers. See adjusted validity/reliability coefficient. [S]	Restriction of range / Restriction of variability	קיצוץ תחום
The result of relating scores on tests. See alternate forms, equating, calibration, moderation, projection, and vertical scaling. [S]	Linking (score linking)	קישור (סולמות מדידה)
In linking test scores for tests that measure similar constructs, the process of relating scores on one test to scores on another, so that scores have the same relative meaning for a group of test takers. [S]	Concordance	קישור ציוני מבחנים
ר		
	Multidimensionality	רב-ממדיות
In classification of disorders, the proportion of cases in which a disorder is detected when it is in fact present. [S]	Sensitivity	רגישות
	Regression	רגרסיה / תסוגה
An interval between two values on a score scale within which, with specified probability, a parameter of interest lies. [S]	Confidence interval	רווח בר-סמך

	Rotation	הוטציה / סיבוב
The granting, usually by a government agency, of an authorization or legal permission to practice an occupation or profession. See also certification, credentialing. [S]	Licensing	רישוי
In testing, variances accruing from the separate constituent sources that are assumed to contribute to the overall variance of observed scores. Such variances, estimated by methods of the analysis of variance, often reflect situation, location, time, test form, rater, and related effects. [S]	Variance components	רכיבי שונות
	Level	רמה / שלב
	cheating	רמייה בבחינה
Brief term or statement indicating a test taker's competency in a particular area of knowledge or skill, usually defined as ordered categories on a continuum, often labeled from "basic" to "advanced," or "novice" to "expert," that constitute broad ranges for classifying performance. See cut score, proficiency level descriptor, and standard setting. [S]	Performance level	רמת ביצוע
	Basal level	רמת בסיס
	Age level	רמת גיל
Descriptions of a test taker's level of competency in a particular area of knowledge or skill, usually defined as ordered categories on a continuum, often labeled from "basic" to "advanced," or "novice" to "expert," that constitute broad ranges for classifying performance. See cut score. [S]	Achievement level / Proficiency level / Mastery level	רמת הישגים / רמת בקיאות / רמת שליטה
	Social desirability	רציות חברתית
	Anecdotal record	רשומת תיאורי התנהגות
	Checklist	רשימת תיג / רשימת בדק
ש		
A questionnaire or checklist, usually in the form of a self-report, that elicits information about an individual's personal opinions, interests, attitudes, preferences, personality characteristics, motivations, and typical reactions to situations and problems. [S]	Inventory	שאלון
An inventory that measures one or more characteristics that are regarded generally as psychological attributes or interpersonal proclivities or skills. [S]	Personality inventory	שאלון אישיות
	Environmental inventory	שאלון משתני סביבה
	Residual	שארית
	Heteroscedasticity	שויון שונות בין קבוצות
A statistic that describes how much the scores in a particular group vary; it is a measure of variability. Some statistical techniques used in testing depend on being able to partition the variance and attribute the parts to various test, test administration, or test taker characteristics. Statistically, it is also the same as the standard deviation squared. [N]	Variance	שונות
	Error score variance	שונות ציוני הטעות
	Reproducibility	שחזוריות
	Facet	שטחה
	Polarizing facet	שטחה מקטבת

	Ordered facet	שטחה סדורה
The actual use of a test, after initial test development has been completed, to inform an interpretation, decision, or action based, in part, upon test scores. [S]	Operational use	שימוש תפעולי
		שינויים במבנה מבחן – ר' התאמות במבנה מבחן
	Base rate	שיעור בסיס
The score(s) obtained by the largest number of individuals in a group. It is the most-frequently-occurring score(s) within a given set of scores. [N]	Mode	שכיח
	Frequency	שכיחות
1. The extent to which the construct measured by one test is essentially the same as the construct measured by another test. 2. The degree to which a construct measured by a test in one cultural or linguistic group is comparable to the construct measured by the same test in a different cultural or linguistic group. [S]	Construct equivalence	שקילות מבנה
In evaluating test translations, the degree to which similar activities or behaviors have the same functions in different cultural or linguistic groups. [S]	Functional equivalence	שקילות פונקציונלית
ת		
1. In test administration, maintaining a consistent testing environment and conducting the test according to detailed rules and specifications, so that testing conditions are the same for all test takers. 2. In test development, establishing scoring norms based on the test performance of a representative sample of individuals with which the test is intended to be used. 3. In statistical analysis, transforming a variable so that its mean is 0.0 and its standard deviation is 1.0 for some specified population or sample. See standard score. [S]	Standardization	תקנון
A kind of standard score for which the digits 1 through 9 are used to describe test performance. The mean value is 5 and the standard deviation is 2. [N]	Stanine	תשעונית
A mathematical model of the relationship between performance on a test item and the test taker's standing on the construct being measured, usually denoted as theta (θ). In the case of items scored 0 / 1 (incorrect/correct response) the model describes the relationship between θ and the item mean score (P) for test takers at level θ , over the range of permissible values of θ . In most applications, the mathematical function relating P to θ is assumed to be a logistic function that closely resembles the cumulative normal distribution. [S]	Item response theory (IRT)	תאוריית התגובה לפריט / תורת התגובה לפריט
	Structuple	תבנה
	Item shell	תבנית פריט

An indicator attached to a test score, a test item, or other entity to indicate a special status. A flagged test score generally signifies a score obtained from a modified test or test administration resulting in a change in the underlying construct measured by the test. A flagged test item generally signifies an item with undesirable characteristics, such as excessive differential item functioning. [S]	Flag	תג
	Ideal response	תגובה אידיאלית
	Dichotomous response	תגובה דיכוטומית / תגובה דו-ערכית
		תואם גיל – ר' מעוגן גיל
		תואם דרגת כיתה – ר' מעוגן דרגת כיתה
	Direct effect	תוצא ישיר
	Indirect effect	תוצא עקיף
The degree to which accumulated evidence and a theory support specific interpretations of test scores for a given use of a test. If multiple interpretations of a test score for different uses are intended, validity evidence for each interpretation is needed. [S]	Validity	תוקף
	Ecological validity	תוקף השלכתי / תוקף אקולוגי
Evidence based on the relationship between test scores and outcomes that are intended to measure different constructs. [S]	Discriminant evidence / Divergent validity	תוקף מבחין
The degree to which evidence supports the proposed interpretation of test scores for a given use. In the current standards, all test scores are viewed as measures of some construct, so the phrase construct validity is redundant with validity. The validity argument establishes the construct validity of a test. See construct, validity argument. [S]	Construct validity	תוקף מבנה
	Incremental validity	תוקף מוסף
Information gathered in the process of validation to show the extent to which scores from one test might be used in place of, or interchangeably with, those from another test. (See also criterion-related evidence and validity.) [N]	Concurrent validity	תוקף מקביל
Evidence based on the relationship between test scores and other measures of the same construct. [S]	Convergent evidence / Convergent validity	תוקף מתכנס
Predictive evidence indicating how accurately test data can predict criterion scores that are obtained at a later time. [S]	Predictive validity	תוקף ניבוי
	Face validity	תוקף נראה
The degree to which evidence based on test content supports the intended interpretation of test scores for a given purpose. Such evidence may address issues such as the fidelity of test content to performance in the domain in question and the degree to which test content representatively samples a domain such as a course curriculum. [S]	Content validity	תוקף תוכן
Information gathered to support the argument that a test does measure the same thing as some other instrument or that it does not measure the same thing as some other particular instrument. Scores from the "other" instrument are referred to as criterion scores. (See also validity.) [N]	Criterion-related validity	תוקף תלוי-קריטריון

An extension of classical reliability theory and methodology in which the magnitudes of errors from specified sources are estimated through the use of one or another experimental design, and the application of the statistical techniques of the analysis of variance. The analysis indicates the generalizability of scores beyond the specific sample of items, persons, and observational conditions that were studied. [S]	Generalizability theory	תורת ההכללה
A psychometric theory based on the view that an individual's observed score on a test is the sum of a true score component for the test taker and an independent measurement error component. [S]	Classical test theory	תורת המבחנים הקלאסית
The construct domain of a variable used as a criterion. See construct domain. [S]	Criterion domain	תחום הקריטריון
The set of interrelated attributes (e.g., behaviors, attitudes, values) that are included under a construct's label. A test typically samples from this construct domain. [S]	Construct domain	תחום מבנה / תחום תכונה
	Item domain	תחום פריט
The set of behaviors, knowledge, skills, abilities, attitudes or other characteristics to be measured by a test, represented in a detailed specification, and often organized into categories by which items are classified. [S]	Content domain	תחום תוכן
		תחזית – ל' ניבוי
Statements that describe what students at a given level of proficiency know and are able to do. [S]	Performance level descriptor	תיאור רמת ביצוע
The body of literature (e.g., test manuals, manual supplements, research reports, publications, user's guides, etc.) made available by publishers and test authors to support test use. [S]	Documentation	תיעוד
In assessment, a systematic collection of educational or work products that have been compiled or accumulated over time, according to a specific set of principles. [S]	Portfolio	תיק עבודות
	Correction for attenuation	תיקון עבור החלשה
	Correction for omission	תיקון עבור השמטה
	Correction for guessing	תיקון עבור ניחוש
The process through which the validity of the proposed interpretation of test scores for their intended uses is investigated. [S]	Validation	תיקוף
A procedure in which a scoring system or set of weights for predicting performance, derived from one sample, is applied to a second sample in order to investigate the stability of prediction of the scoring system or weights. [S]	Cross-validation	תיקוף צולב
	Attribute	תכונה / מאפיין
Knowledge, skills, abilities, or traits that an assessment is intended to measure. Constructs can include such concepts as study skills, honesty, knowledge of algebra, or the ability to drive a car. [S]	Construct	תכונה / מבנה / קונסטרוקט
	Latent trait	תכונה חבויה

An approach to assessment development that attempts to maximize the accessibility of a test for all of its intended test takers. When using Universal Design, test developers differentiate what is both relevant and irrelevant to the intended construct and test purpose. Test items and tasks can then be purposively designed and developed to address the construct(s) to be measured and to minimize construct-irrelevant features that might otherwise impede the performance of intended test taker groups, for example, individuals with disabilities, those from diverse linguistic or cultural groups, older adults, or young children. [S]	Universal design	תכנון מותאם לכול
The activity of a practitioner that involves the development of a treatment protocol. [S]	Intervention planning	תכנית התערבות
		תנאים מותאמים – ר' התאמות
		תפקוד – ר' פונקציה
Differential test functioning (DTF) is a term similar to DIF but used to describe the test or dimension level. When DTF occurs, individuals from different groups who have the same standing on the characteristic assessed by a test do not have the same expected test score. [S]	Differential test functioning	תפקוד דיפרנציאלי של מבחן
A statistical indicator that different groups of test takers who have the same total test score have different average item scores or, in some cases, different rates of choosing various item options. Also known as DIF. [S]	Differential item functioning	תפקוד דיפרנציאלי של פריט
The ability of an individual to perform the various mental activities most closely associated with learning and problem solving. Examples include verbal, spatial, psychomotor, and processing-speed ability. [N]	Cognitive functioning	תפקוד קוגניטיבי
	Observation	תצפית
A standard, number .0, that serves as a guiding principle for all of the standards in a chapter. [S]	Overarching standard	תקן על
1. An objective definition of a certain level of performance in some domain in terms of a cut score or a range of scores on the score scale of a test measuring proficiency in that domain; established using a standard setting method. 2. A statement or description of a set of operational tasks exemplifying a level of performance associated with a more general content standard; the statement may be used to guide judgments about the location of a cut score on a score scale. The term often implies a desired level of performance. See cut score. [S]	Performance standard	תקן של ביצוע

In educational assessment, a statement of content and skills that students are to learn in a subject matter area at a particular grade or at the completion of a particular level of schooling. [S]	Content standard / Content standard	תקן תוכן
Alternate or alternative standards are terms used in educational assessment to denote content and performance standards for students with significant cognitive disabilities. Alternative standards are intended to align with regular content standards by extending the lower end of the continuum of performance expectations, ranging from very rudimentary to approaching grade level. [S]	Alternate standards / Alternative standards	תקנים חלופיים
A translation of a test, which is itself a translation from an original test, back into the language of the original test. The degree to which a back translation matches the original test indicates the accuracy of the original translation. [S]	Back translation of a test	תרגום חזרה של מבחן