

Definitions of terms used in the European Test User Standards

European Test User Standards

for test use in Work and Organizational settings

Document 2: GLOSSARY OF TERMS¹

A

ABILITY

ABILITY describes the degree to which someone can carry out certain types of mental operations -- generally operations which involve 'reasoning' of some form.

ABILITY MEASURE *see* ABILITY TEST

ABILITY TEST

ABILITY TESTS vary in the types of operation they involve and the types of material they contain. Typical operations include 'analogies' (A is to B as C is to ?) and 'series completion' (1,2,4,8,?). The content of ability test ITEMS tends to concern words and sentences, numbers or shapes. Ability tests are generally designed to assess what people are capable of rather than what they have learnt or what they know.

ACCEPTABILITY

ACCEPTABILITY concerns those factors which make a test acceptable to the test taker and the test user. The notion of ACCEPTABILITY is important in the design and administration of assessment procedures and should be taken into account to ensure that the test taker cooperates in the procedure. ACCEPTABILITY is affected by the FACE VALIDITY of the test and by the test user's faith in it.

ACCURACY

The precision with which some attribute is measured. ACCURACY is a function of RELIABILITY or freedom from measurement error.

ACHIEVEMENT TEST *see* ATTAINMENT TEST

¹ These definitions have been taken either from Bartram & Lindley (1994, 2000), *Psychological Testing: The BPS Level A Open Learning Programme* or from Bartram & Lindley (2005). *Psychological Testing: The BPS Test Administration (Occupational) Open Learning Programme*, both published by BPS Blackwell. The contents of this Glossary are copyright © David Bartram and Patricia Lindley (1994, 2000, 2005) and are reproduced here by permission.

ADAPTIVE TESTING

see ADAPTIVE TEST.

ADAPTIVE TEST

A TEST where the items are selected from a large database or bank of items held on the computer. Each person who takes the test may be given a different selection of items, as the computer picks just those items that provide most information about that particular person's level of ability.

ADMINISTRATOR

The person who administers a psychological TEST.

ADVERSE IMPACT

The effect of a selection process which results in relatively more people from one group being selected than from another. ADVERSE IMPACT is one, but only one, of the conditions required to establish indirect discrimination. Also known as disparate or disproportionate impact.

APPARATUS TEST

A TEST that requires the manipulation of various items of specialised apparatus -- peg-boards, typing tests, etc.

APPLICANT

A person who has made an application for a position.

APTITUDE BATTERY

A sequence of APTITUDE TESTS which can be used to provide both detailed and general overall measures.

APTITUDE TEST

An APTITUDE is a potential to succeed at something in particular. ABILITY is assumed to underlie aptitude. APTITUDE TESTS are those which have been designed to measure those mental operations (or abilities) which affect the likelihood of someone acquiring some particular skill (for example, computer programming or TV repairing). ABILITY TESTS differ from aptitude tests in that the former are designed to assess the general reasoning skills while the latter tend to contain ITEMS with content which more specifically related to the aptitude concerned. However, the difference is largely one of function or use. In many cases, the same actual test may be used either as an ability test (to measure a person's general

Definitions of terms used in the European Test User Standards

intellectual functioning) or an aptitude test (to assess their potential for success in some occupation). In practice, measures of GENERAL ABILITY can usually be drawn from aptitude test scores (especially APTITUDE BATTERIES).

ASSESSMENT

The process of appraising or estimating some attribute or set of attributes of a person.

ATTAINMENT TEST

The focus of attainment or achievement TESTS is knowledge and proficiency; on what has been learnt rather than on the ability to learn. These tests specifically assess what people have learnt and the skills they have acquired, for example, shorthand and typing tests.

AUTHENTICATION

One of the functions of test administration is to authenticate the identity of test takers. That is, to ensure that the person who presents themselves for testing is actually who they say they are. For high stakes testing, the test administrator should ask for some form of identification to confirm this. This should be some form of photo identification, not something that might have been given to an accomplice.

B

BIAS

BIAS occurs in TESTS whenever people's responses vary in some systematic way which is related to some characteristic which the test was not intended to measure. Factors which can produce bias in scores (at either the item level or overall test score level) include differences in sex, age, culture, educational background and literacy.

BREADTH

Assessment methods vary in both their BREADTH and their SPECIFICITY. A TEST of GENERAL ABILITY that samples several ABILITY DOMAINS may be regarded as `broad' and `general'.

C

CANDIDATE

A person who has taken part in the SELECTION process for a particular job or training course.

CLASSICAL TEST THEORY

A theoretical approach used in PSYCHOMETRICS that regards all observed scores as fallible and defines the relationship between observed scores and the TRUE SCOREs which are assumed to underlie them.

CLIENT

The person or persons who request a service (in this case, testing) either for themselves or as a representative of an organization.

COGNITION

The internal processes and operations involved in perception, memory, thinking, reasoning and problem solving.

COMPETENCIES

Work-related behaviours that are instrumental in the delivery of desired result or work outcomes.

COMPETENCY PROFILE

A profile or list of the COMPETENCIES required for a job or position.

COMPOSITE TEST SCORE

A score produced by adding together scores of two or more tests or SUB-TESTS. In some cases, these may be differentially weighted before they are added. COMPOSITE TEST SCORES are frequently produced by simply summing the RAW SCORES for each part of a TEST. COMPOSITE scores derived from BATTERIES of tests are often produced to provide a general measure of suitability in a selection situation.

CONFIDENTIALITY

Any information obtained using a psychological test should be considered as belonging to the test candidate in the first instance. Whatever is done with that information should be done only with their INFORMED CONSENT.

CONTROLLED MODE

This is a mode of test administration in which control is exercised over who can access a test on the internet and how often they can access it. It may also include controls over the location they can access it from and the time or date it is available.

CRITERION REFERENCING

In CRITERION REFERENCING a person's score on a TEST is used to predict or anticipate how they will perform on types of task not directly sampled by the test but which have been shown to be correlated with test performance.

D

DATA PROTECTION and DATA PRIVACY LAWS

Legislation that gives individuals rights concerning personal data that is stored on computer or in other filing systems.

DECILE

A percentile-based scoring system where the RAW SCORES are divided into ten categories each containing 10% of the distribution.

DEXTERITY TEST

DEXTERITY TESTS are designed to assess various aspects of motor coordination, such as speed of movement, precision of fine motor control etc.

DISPOSITION

The temperament, personality, or characteristic mode of operating of a person.

DISTRIBUTION

In psychometric testing terms, the apportionment of scores obtained by people across all possible values of a variable. *See* FREQUENCY DISTRIBUTION.

DOMAIN

A universe, sphere or province of objects which meet some criteria. In relation to psychological testing people talk about ability TEST ITEMS being drawn, for example, from the DOMAIN of Verbal Reasoning, or from the domain of Spatial Ability. The term is also applied to areas of achievement. For example, mathematical attainment is a domain

Definitions of terms used in the European Test User Standards containing a range of skills involving arithmetic, algebra and so on.

DOMAIN-REFERENCED MEASURE

This is where CONTENT-RELATED VALIDITY data is used as the basis for interpreting a test score. The logic is to relate performance on a TEST to the level of performance required in a job by using a common quality standard. Judgements are then made about what level of performance on the test would be required for adequate performance of the job. The weakness of domain-referencing is that it relies on expert judgement. As such it really provides a way of generating hypotheses about how test performance should relate to job performance: it does not actually prove that the two are related. Where possible it needs to be backed up by CRITERION-RELATED VALIDITY studies.

DOMAIN REFERENCING

The process of relating a person's score on a TEST to levels of competence within some DOMAIN of knowledge or performance.

DRIVE

An attribute or need of a person which is considered to cause them to act in a certain way or motivate them to action.

E

ETHICAL (ISSUES)

Issues concerned with the rights, responsibilities and obligations of those involved in testing -- the test taker, the test user, and the test user's client.

F

FACE VALIDITY

What, to the test taker, the TEST appears to measure. The superficial appropriateness of a test (see ACCEPTABILITY).

FAIRNESS

Fairness in testing is a relative term. A TEST is fair or unfair depending on whom it is used and how it is used. Its use is fair if it is not BIASED with respect to the groups with which it

Definitions of terms used in the European Test User Standards

is used and if it can be shown to be valid. Thus, using either an unbiased test of mechanical reasoning or a biased clerical aptitude test in a clerical selection situation would be unfair. In both cases, differences between people would not be relevant for the job and hence it would be unfair to select people on the basis of such differences. Where a test is known to be valid but shows between group bias and the degree and type of test BIAS is known with respect to some selection procedure, then it is possible to practise fair selection by using different cut-off scores for each group.

FAIR SELECTION

The application of the principles of FAIRNESS to the selection process.

FEEDBACK *see* REPORTING BACK

FIVE POINT GRADING SCHEME

A common percentile-based scoring system where the top 10% of scores are classed as grade A; the next 20% as grade B; the next 40% as grade C; the next 20% as grade D and the lowest 10% as grade E.

FREQUENCY DISTRIBUTION

The number of people who obtained each of the various values which could be obtained on a particular VARIABLE. A FREQUENCY DISTRIBUTION shows how people's scores are distributed across all possible values. FREQUENCY DISTRIBUTIONS are often used to examine the number of people obtaining each of the possible RAW SCORES on a TEST.

G

GENERAL ABILITY

ABILITY TESTS vary from those designed to give an overall measure of general intellectual functioning (GENERAL ABILITY TESTS) through those designed to assess broad areas of ability (for example, Verbal, Numerical or Spatial) to those focusing on specific MENTAL OPERATIONS (for example, three-dimensional spatial rotations). The latter tend to be used for aptitude assessment. General ability tests, in order to properly cover the full range of mental operations, tend to include ITEMS or SUBTESTS dealing with each of the main areas of ability. When GENERAL ABILITY is tested using a battery of ability TESTS, SPECIFIC ABILITY scores as well as an overall general ability measure can be obtained.

GENERAL ABILITY TEST

A TEST designed to give an overall measure of general intellectual functioning. *See* GENERAL ABILITY.

GENERAL INTELLIGENCE

The ability to perform on TESTS and in tasks which involve the understanding of relationships. The capacity to meet new situations, or to learn to do so by new adaptive responses.

GENERAL NORMS

General norms are intended to be representative of a large and diverse POPULATION. For example, UK general population; Norwegian adult males; French 16 to 18 year old school leavers. *See* NORMS and SPECIFIC NORMS.

GENERAL POPULATION NORMS

NORMS suitable for use with most people for converting RAW SCORES to either PERCENTILE or STANDARD SCORE based measures. Such norms are usually based on a large representative sample of people (in terms of age, sex and other VARIABLES).

GENERALISABILITY THEORY

This theory is concerned with how much we can generalize from scores obtained under one set of conditions to those we would be likely to obtain under another. This approach attempts to measure the effects of separate sources of ERROR VARIANCE which can affect the RELIABILITY of a TEST and then estimate the amount of ERROR that will be present in measures obtained under various different sets of conditions.

GENERATING REPORTS

The process of creating a report on a person's test results by using computer software. Computer-generated reports may be obtained either from stand alone PC applications or from service providers on the Internet

GRADES *See* FIVE POINT GRADING SCHEME

GUIDANCE

The process of giving information to a person to help initiate, support and clarify their decisions.

H

HISTOGRAM

A FREQUENCY DISTRIBUTION represented in a graphical form. Each score is represented by a bar, the height of which is equal to the number of people who obtained that score.

I

IDENTIFICATION OF DEVELOPMENT NEED

The use of tests to identify those areas of behaviour where a person may have potential for change. See PERSONAL DEVELOPMENT.

INFORMED CONSENT

The gaining of agreement from a test taker to the terms and conditions under which testing is to take place. This agreement must be obtained having first clearly informed the test taker of their rights and responsibilities, the reasons for testing, the type of tests to be used and what will be done with the results of the tests.

INSTRUMENT

A psychological TEST or other procedure for measuring differences between people.

INTEREST INVENTORY

An INTEREST INVENTORY is designed to assess, in a systematic manner, people's likes and dislikes for different types of work or leisure activity.

INTERESTS

Attitudes towards various types of activity, either in relation to work (vocational interests) or outside work.

IPSATIVE TEST

An IPSATIVE TEST compares a person's score on one scale with their own score on other scale(s). As a result, the scores on each scale are dependent on each other to some degree. Sometimes referred to as a self-referenced TEST.

ITC

The International Test Commission. This is a body that sets Guidelines relating to tests and test use through international consultation. Further details are on the ITC website www.intestcom.org

ITEM RESPONSE THEORY

A theory defining the relationship between test items and the likelihood of people making correct responses to those items in terms of the level of the 'latent trait' (ability) they possess. The theory forms the basis for many ADAPTIVE TESTING systems.

ITEM SCORE

The numerical score given to a test taker's answer to an individual TEST item. For example, a correct response to an item in a numerical reasoning test may be scored as 1 and an incorrect response may be scored as 0.

ITEM

An ITEM is the smallest element within a test to which a score is assigned. Generally a question within a TEST, or a statement in a personality questionnaire.

J

JOB ANALYSIS

The process of formally defining the nature of a job, often in terms of the characteristics required by people who are to perform it.

JOB DESCRIPTION

A description of a job produced from the results of a JOB ANALYSIS

JOB-RELATED KNOWLEDGE ATTAINMENT TEST

A TEST designed to measure what a person knows in relation to a particular job.

JOB SIMULATION

JOB SIMULATION exercises are often used in the procedures that come under the general heading of the Assessment Centre Method. Job simulations may take the form of in-tray exercises, group problem-solving exercises and so on. They start from the assumption that the candidate does not yet possess the requisite knowledge or skill, but that the underlying ability will manifest itself when he or she works through an exercise that simulates the broad

demands of the job in question. Job simulations are generally designed to assess APTITUDE rather than ATTAINMENT, though they may rely on the acquisition of some general job competences.

L

LEVEL OF DIFFICULTY

The level of difficulty concerns the degree of ABILITY required to answer a test ITEM. TESTS may either be designed to contain items of a similar difficulty level, or the level of difficulty may be increased as the test taker progresses through the test. The idea behind this approach is that it should provide a wider range of discrimination between people with the more able people getting further into the test.

LOCAL NORM

A particular type of specific NORM GROUP. This is a sample of people local to an organization. *Also see* NORMS.

M

MANAGED MODE

A mode of administration in which there is both direct supervision and control over the equipment being used, and other conditions. Typically managed mode administration refers to the use of dedicated testing centres.

MAXIMUM PERFORMANCE

Measures of MAXIMUM PERFORMANCE measure how well people can do things, how much they know and how great their potential is. Measures of maximum performance include TESTS of ABILITY, APTITUDE and ATTAINMENT. These measures are usually distinguished from measures of TYPICAL PERFORMANCE which assess personality, vocational or occupational interests, needs, drives and levels of motivation.

MEASUREMENT

The procedure used to obtain a score from a person on some SCALE.

MEASUREMENT ERROR

Inaccuracies arising from the measurement process, the sources of which are due to extraneous factors which affect scores on a TEST. *See* SYSTEMATIC BIAS and RANDOM ERROR.

MENTAL OPERATIONS

MENTAL OPERATIONS are the internal processes and manipulations that have to be carried out to answer a TEST ITEM.

MODES OF TEST ADMINISTRATION

Classification by the International Test Commission of test administration modes into four types: OPEN, CONTROLLED, SUPERVISED and MANAGED.

MOOD INVENTORY

An instrument designed to assess a person's mood.

MOTIVATION

Factors which affect a persons likelihood of action and the choices they make between alternative courses of action. The reasons given as to why people perform certain acts. Motivation is often defined in terms of the goals which people seek to attain through their actions.

MULTIPLE-CHOICE (ITEM FORMAT)

For MULTIPLE-CHOICE items, test takers have to select one of a number of possible answers. Also see OPEN-ENDED (ITEM FORMAT).

N

NEED

A requirement which can act as the driving force or motivation for action. Needs include nurturance, affiliation, social approval, sex, self-actualization etc.

NORM GROUP

The sample of people from whom NORMS are derived. Also referred to as a REFERENCE GROUP.

NORM TABLES *See* NORMS

NORMAL CURVE *See* NORMAL DISTRIBUTION

NORMAL DISTRIBUTION

A symmetrical bell-shaped distribution with certain specific properties: the MEAN, MODE and MEDIAN are all equal to each other; the proportion of the values falling between any interval along the scale is known from the mathematical properties of the distribution. There will always be, for example, 68% of the values between -1 and +1 STANDARD DEVIATIONS. This form of distribution is found for a wide variety of both physical and psychological traits. Also called the 'normal frequency distribution' or the 'normal curve'.

NORMATIVE

Normative information included in a TEST's DOCUMENTATION enables the test user to see how a person's performance on the test compares with that of others.

NORMATIVE SCORE *See* NORM-REFERENCED MEASURE

NORM-REFERENCED MEASURE

A NORM-REFERENCED MEASURE defines where a person's raw score lies in relation to the scores obtained by other people (that is, a NORM GROUP).

NORM-REFERENCED SCORE

The score obtained on a NORM-REFERENCED MEASURE. Such scores are expressed either as some form of PERCENTILE SCORE or STANDARD SCORE.

NORMS

Information usually in the form of a table, which enables RAW SCORES to be converted into PERCENTILE SCORES or STANDARD SCORES (or both). Also see NORMATIVE, GENERAL NORMS and SPECIFIC NORMS.

O

OPEN-ENDED (ITEM FORMAT)

For OPEN-ENDED ITEMS, test takers have to write down their own responses to the items: Alternatives are not given. *See also* MULTIPLE CHOICE (ITEM FORMAT)

OPEN MODE

Open Mode is where the test taker has direct access to the test materials. So there is no involvement of a test user or test administrator. Such tests include the books of tests you might buy in the local book shop or the tests you can find on the Internet that are directly accessible to everyone. Often the only requirement is that you pay some money before you can access the test. However, no qualifications are required from you either in terms of test use or test administration.

ORAL FEEDBACK *See* REPORTING BACK

P

PERCENTILE *See* PERCENTILE RANK and PERCENTILE SCORE

PERCENTILE RANK

The value on the RAW SCORE scale below which a given percentage of the sample's scores lie. For example, if the 85th percentile rank is 16, then 85% of the sample will have scored less than 16. The PERCENTILE RANK is more commonly referred to as just the PERCENTILE or in some cases the CENTILE.

PERCENTILE SCORE

A number between 0 and 100 expressing a test taker's RAW SCORE in terms of the percentage of the norm group who scored less. *Also see* PERCENTILE RANK.

PERFORMANCE MANAGEMENT

Procedures used in organisations to help in managing the performance of employees towards the achievement of organisational goals and objectives.

PERSONAL DEVELOPMENT

Procedures by which individuals can develop their knowledge skills and COMPETENCIES. Testing may be used to help identify where their strengths lie and where they may have development needs.

PERSONALITY INVENTORY

Psychological TESTS that assess DISPOSITION, that is, preferred or typical ways of acting or thinking. Personality inventories attempt to measure how much or how little a person possesses of a specified TRAIT or set of traits.

PERSON SPECIFICATION

A description derived from a **JOB DESCRIPTION** of the personal characteristics necessary to do that job.

PHYSICAL MAKE-UP

For example, health, physique, appearance, bearing and speech.

POPULATION

A **POPULATION** contains all the people who conform to some specification. In **PSYCHOMETRICS**, normative reference groups are populations: for example, UK adult females; university arts graduates; general population. Psychometrics involves making inferences about people who come from some population on the basis of information known about the behaviour of a representative sample from that population.

POTENTIAL

A capacity or capability to perform or acquire the skills to perform some class of actions.

POWER TEST

The focus of a **POWER TEST** is on how many items a person is able to answer correctly. The time limit is designed to allow most people to complete all of the test **ITEMS**. If a person's score is mainly affected by their ability to answer the questions correctly -- rather than their speed -- the test is a power test.

PRACTICALITY

The notion of **PRACTICALITY** in **TEST** use concerns issues of cost-efficiency, such as what the test costs; what training is required to use it; how long it takes to administer, score and interpret; what equipment is needed.

PRIVACY

Ensuring that people's test data and other personal information is not disclosed to others without the **INFORMED CONSENT** of the test taker.

PROFESSIONAL (ISSUES)

Issues associated with professional practice and codes of professional conduct in relation to test use -- see also **ETHICAL ISSUES**.

PSYCHOLOGICAL TEST *See* **TEST**

PSYCHOLOGICAL TESTING

The use of psychological TESTS in the process of assessment.

PSYCHOMETRICS

Literally, the measurement of mental processes. PSYCHOMETRICS is the technology that underlies TESTS and their development.

Q

QUARTILE

A percentile-based scoring system where the RAW SCORES are divided up into four categories each containing 25% of the distribution (referred to as the first, second, third and fourth QUARTILES).

R

RAW SCORE

The RAW SCORE is the total number of correct answers a person obtains on completing a TEST. A raw score is the ABSOLUTE score a person gets on a test.

RECRUITMENT

The process of obtaining new hires into an organisation. In some cases, the term RECRUITMENT is used to refer to the whole process from applicant attraction through to hire. Others may use it to refer to the processes leading up to the definition of a short list from which final choices are made (SELECTION).

REFERENCE GROUP *See* NORM GROUP

REFERENCING SCORES

To compare a person's RAW SCORES on a SCALE against some other measure. The comparison may be with other people's scores on the same scale, the person's own scores on other scales, known relationships with other performance measures, expected levels of attainment in the domains from which the test items were drawn. *See* NORM-REFERENCED, SELF-REFERENCED, CRITERION-REFERENCED and DOMAIN-

REFERENCED.

RELATIONSHIPS

TEST ITEMS where a test taker has to identify the relationship between two or more things and then use that relationship to select from a set of alternatives.

RELATIVE SCORE

A **SCALE SCORE** which describe a person's performance relative to that of other people or in terms of some other measure.

RELEVANCE

Another word used to describe the concept of **VALIDITY**.

RELIABILITY

The extent to which one can rely on the obtained **TEST** score being an accurate measure of a person's **TRUE SCORE**, rather than a measure of incidental random factors. **RELIABILITY** is usually assessed either by measures of **INTERNAL CONSISTENCY**, **EQUIVALENCE** or **STABILITY**.

REPORTING BACK

The process of feeding back the **INTERPRETATION** of **TEST** scores to a **CLIENT** or **CANDIDATE**. This can be in the form of oral feedback or a written report.

S

SAMPLING

The selection of a limited number of people (or other objects) from a defined **POPULATION**.

SCALE

In testing it is common to talk of measuring characteristics along a scale. Ability, for example, being a scale which goes from low to high scores. Thus scores obtained on a **TEST** of some characteristic are generally referred to as **SCALE SCORES**. *See* **SCALE SCORES**, **RAW SCALE SCORE**.

SCALE SCORE

The numerical scores attributed to a test taker's answers to individual **TEST** items are added up to provide a single measure called a **RAW SCALE SCORE**. Scores obtained on a test are

Definitions of terms used in the European Test User Standards generally referred to as SCALE SCORES.

SCOPE

In relation to psychological TESTS, the range of attributes covered by a test (the test DOMAIN) and the range of people with whom the test can be used.

SCORING

The process of marking the answers to a psychological TEST, including conversion of RAW SCORES into STANDARD SCORES.

SCORING KEYS

Tools for assisting in the marking of tests. For example, many scoring keys are in the form of templates that can be laid over the answer sheets.

SECURITY (OF TEST MATERIALS)

Guarding against the unfair and illegal use of TEST MATERIALS.

SELECTION

The process of choosing people with the best chances of succeeding in a job or on a training course.

SELF-REFERENCED MEASURES

SELF-REFERENCED MEASURES involve comparing a person's scores on one scale with their scores on other scales. *See* IPSATIVE TEST.

SELF-REPORT

SELF-REPORT measures are instruments that ask the respondent to answer a structured set of questions about themselves. Also called self-description instruments. Most personality and interest TESTS are self-report.

SPECIAL APTITUDE

Another way of describing a SPECIFIC APTITUDE, for example, mechanical, dexterity and so on.

SPECIFIC ABILITY

A particular ABILITY (for example, spatial ability).

SPECIFIC APTITUDE

A particular aptitude (for example, clerical accuracy).

SPECIFICITY

The degree to which a test assesses specific as opposed to general ability or aptitude.

SPECIFIC NORMS

NORMS which are based on some specific sample. For example, public sector engineering workers; clerical staff from a number of different companies. Also see NORMS and GENERAL NORMS.

SPEED TEST

A TEST which contains relatively easy ITEMS but which have a strict time limit. The measure of performance stresses the number of items attempted within the fixed time. *See* POWER TEST.

SPEEDED

It is common to see references to TESTS as more or less speeded. The more the standard time limit results in people failing to attempt some items, the more the test is said to be 'speeded'. *See* SPEED TEST and POWER TEST.

STANDARD SCORE

In PSYCHOMETRICS, the 'standard' scale developed for measuring psychological characteristics is called the Z-SCORE SCALE or sometimes simply the STANDARD SCORE SCALE. A z-score is a measure equal to one STANDARD DEVIATION (SD) of a distribution.

STANDARD SCORE SCALE

The most commonly used standard score scales are z-scores (one SD), T-scores (one tenth of an SD), STENS (half an SD), STANINES (half an SD) and IQ (generally one fifteenth or one twentieth of an SD depending on the scale used). *See also* STANDARD SCORES.

STANDARD ERROR OF MEASUREMENT (SE_m)

The amount of ERROR associated with making inferences about a person's TRUE SCORE from their obtained score.

STANDARDISATION

The procedure of establishing the initial set of NORMS for a TEST, defining the conditions under which it should be used, and of assessing its RELIABILITY and VALIDITY.

STANDARDISED

In PSYCHOMETRICS, a standardised measure, TEST or testing procedure is one which has known characteristics. For example, a CORRELATION COEFFICIENT is a standardised score (like a Z-SCORE) which is known to come from a distribution of possible values which range between zero and plus or minus one.

STATES

STATES are concerned with how a person is feeling or performing at a particular moment in time (for example, a current mood), rather than how they generally feel or typically perform. States are often distinguished from TRAITS which are more stable and enduring psychological characteristics.

SUPERVISED MODE

This is the mode in which the test administrator has direct face-to-face involvement with the test taker. The test takers will come to a location where the test administrator is able to supervise them taking the test.

SYSTEMATIC BIAS

A source of MEASUREMENT ERROR which is predictable and can lead to possible unfair BIAS in the use of TESTS. SYSTEMATIC BIAS is potentially measurable.

T

TECHNICAL MANUAL The part of a TEST MANUAL that covers all the technical details of the test, such as design and development, RELIABILITY, VALIDITY, BIAS and STANDARDISATION

TEST

An assessment procedure designed to provide objective measures of one or more psychological characteristics. These include ABILITIES, APTITUDES, ATTAINMENTS, INTERESTS, beliefs, personality and so on. The important feature of psychological TESTS is that they produce measures obtained under standardised assessment conditions which have known RELIABILITY and VALIDITY. They provide a way of comparing a person's

Definitions of terms used in the European Test User Standards performance against that of others. An instrument that has been developed using psychometric principles. The term, test, is used as shorthand for psychological test or psychometric test, which includes various inventories and questionnaires.

TEST ADMINISTRATION

The process of administering a psychological TEST to one or more people.

TEST BATTERY

A sequence of TESTS. *Also see* APTITUDE BATTERY.

TEST DATA

The information about candidates and their scores resulting from the taking of a psychological TEST.

TEST INTERPRETATION

The process of attributing a meaning to a test taker's score on a psychological TEST, by reference to information about the test's validity, reliability and its NORMS.

TEST LENGTH

The number of ITEMS a TEST contains.

TEST MANUAL

The technical documentation accompanying a psychological TEST which tells the test user how to use and administer the test and what conclusions can be drawn from the results. Various referred to as TECHNICAL MANUAL, USER MANUAL or TEST DOCUMENTATION.

TEST MATERIALS

The materials needed to administer and interpret a psychological TEST. These generally include a TEST manual, answer sheets, question booklets, profile sheets, administration instructions and so on.

TEST RELIABILITY

The RELIABILITY of a TEST.

TEST SOPHISTICATION

A level of awareness and knowledge of TESTS or testing without which a person's scores may be negatively biased. TEST SOPHISTICATION may arise from prior exposure, the

Definitions of terms used in the European Test User Standards

process of testing, or through the use of practice tests and information describing testing procedures. This should be distinguished from active coaching in how to do a particular test which will unfairly inflate a person's scores.

TEST SESSION

The period during which a psychological TEST is administered.

TRAINABILITY TEST

A TRAINABILITY TEST is designed to see whether a person is likely to be able to cope with the training required to do a job. Typically, it consists of a highly-structured short training course with a test of performance at the end.

TRAIT

TRAITS are those relatively stable and enduring characteristics of people that make them predictable. Traits are usually distinguished from STATES which are a more changeable form of psychological characteristic, for example, current moods and feelings.

TYPICAL PERFORMANCE

Measures of TYPICAL PERFORMANCE are designed to assess disposition, such as personality, beliefs, values and interests and to measure motivation or 'drive'. Measures of typical performance are usually distinguished from measures of MAXIMUM PERFORMANCE which are designed to assess how well people can do things and measure ability, aptitudes or attainment.

U

UNBIASED

Unbiased tests are those that do not show any systematic difference with regard to particular groups of people. BIAS is always relative to the groups concerned. A test might be unbiased with respect to age, while biased with respect to gender, for example.

USER MANUAL *See* TEST MANUAL

UTILITY

UTILITY concerns the benefits which accrue from the use of a psychological TEST. Utility is a function of the balance between issues of PRACTICALITY (the 'costs' associated with using a test) and RELEVANCE and FAIRNESS (the 'benefits' associated with the test). The

Definitions of terms used in the European Test User Standards benefits, in turn, are limited by the RELIABILITY of the test and its SCOPE.

V

VALIDATION

The process of building up evidence about what can and cannot be inferred from TEST SCORES.

VALIDATION OF RESULTS

In the context of test administration, this refers to the need to ensure that the results have been obtained under proper conditions and have not been the outcome of cheating and collusion.

VALIDATION STUDIES

Research studies that provide evidence on the VALIDITY of a test.

VALIDITY

Information on the validity of a TEST tells the user what inferences can be drawn about the person who has produced the score on a test and what is being measured by a test -- that is, what is it a test of. *See* CONTENT, CONSTRUCT and CRITERION-RELATED VALIDITY.

VARIABLE

VARIABLES are the characteristics which we attempt to measure with psychological TESTS. They are so named because they vary from person to person, or for the same person from time to time. The defining characteristic of a variable is that a given person can have only one value of it at any one time -- for example, they cannot be 5' 5" AND 5' 10" at the same time. There are three main classes of variable: NOMINAL, ORDINAL, and SCALAR.

W

WORK SAMPLE

A **WORK SAMPLE TEST** is one in which the task has been taken from a job. The task is done under **STANDARDISED** assessment conditions. Work samples are essentially **ATTAINMENT TESTS**. They pre-suppose that the test taker has acquired some measure of a particular skill and sets out to see how much.

WORKPLACE COMPETENCIES

See **COMPETENCIES**

Z

Z-SCORE

A **STANDARD SCORE** scale with a **MEAN** of zero and a **STANDARD DEVIATION** of one.