

## Glossary of Terms

- Ability** A defined domain of cognitive, perceptual, psychomotor, or physical functioning.
- Accommodation** A change in the content, format, and/or administration of a selection procedure made to eliminate an irrelevant source of score variance resulting from a test taker's disability.
- Adjusted validity/reliability coefficient** A validity or reliability coefficient—most often a product-moment correlation—that has been adjusted to offset effects of differences in score variability, criterion variability, or unreliability of test and/or criterion. See *Restriction of range or variability*.
- Alternate forms** Two or more versions of a selection procedure that are considered interchangeable in that they measure the same constructs in the same ways, are intended for the same purposes, and are administered using the same directions. *Alternate forms* is a generic term used to refer to either parallel forms or equivalent forms. *Parallel forms* have equal raw score means, equal standard deviations, equal error structures, and equal correlations with other measures for any given population. *Equivalent forms* do not have the statistical similarity of parallel forms, but the dissimilarities in raw score statistics are compensated for in the conversions to derived scores or in form-specific norm tables.
- Analysis of work** Any method used to gain an understanding of the work behaviors and activities required, or the worker requirements (e.g., knowledge, skills, abilities, and other personal characteristics), and the context or environment in which an organization and individual may operate.
- Assessment** Any systematic method of obtaining information from tests and other sources; used to draw inferences about characteristics of people.
- Battery** A set of selection procedures administered as a unit.
- Bias** In a statistical context, a systematic error in a score. In discussing fairness, bias refers to variance due to contamination or deficiency that differentially affects the scores of different groups of individuals.
- Compensatory model** Two or more individual selection procedure component scores (often individual test scores) combined into a composite selection procedure according to some specified formula (including simple summation of scores and unit weighting). As a consequence of combining scores, some compensation for one or more of the constructs measured may occur due to differential performance on the individual selection procedures (i.e., a higher score on one test compensating for a lower score on another test).
- Composite score** A score that combines scores from several individual selection procedures according to a specified formula.

- Concurrent validity evidence** Demonstration of the relationship between job performance and other work outcomes, and scores on selection procedures obtained at approximately the same time.
- Confidence interval** An interval between two values on a score scale within which, with specified probability, a score or parameter of interest is expected to lie.
- Configural scoring rule (Configural scoring)** A rule for scoring a set of two or more elements (such as items or subtests) in which the score depends on a particular pattern of responses to the elements.
- Consequence-based evidence** Evidence that consequences of selection procedure use are consistent with the intended meaning or interpretation of the selection procedure.
- Construct** A concept or characteristic of individuals inferred from empirical evidence or theory.
- Construct irrelevance** The extent to which scores on a predictor are influenced by factors that are irrelevant to the construct. Such extraneous factors distort the meaning of scores from what is implied in the proposed interpretation.
- Contamination** Systematic variance that is irrelevant to the intended meaning of the measure.
- Content domain** A body of knowledge and/or set of tasks, activities, or other personal characteristics defined so that given knowledge, activities, or characteristics may be classified as included or excluded.
- Content-based validity evidence** Demonstration of the extent to which content on a selection procedure is a representative sample of work-related personal characteristics, work performance or other work activities or outcomes.
- Convergent validity evidence** Evidence of a relationship between measures intended to represent the same construct.
- Correlation** The degree to which two sets of measures vary together.
- Criterion** A measure of work performance or behavior, such as productivity, accident rate, absenteeism, tenure, reject rate, training score, and supervisory ratings of job relevant behaviors, tasks or activities.
- Criterion-related validity evidence** Demonstration of a statistical relationship between scores on a predictor and scores on a criterion measure.
- Criterion relevance** The extent to which a criterion measure reflects important work performance dimensions or other work outcomes.
- Critical score** A specified point in a distribution of scores at or above which candidates are considered successful in the selection process. The critical score differs from cutoff score in that a critical score is by definition criterion referenced (i.e., the critical score is related to a minimally acceptable criterion) and is the same for all applicant groups.

- Cross-validation** The application of a scoring system or set of weights empirically derived in one sample to a different sample from the same population to investigate the stability of relationships based on the original weights.
- Cutoff score** A score at or above which applicants are selected for further consideration in the selection process. The cutoff score may be established on the basis of a number of considerations (e.g., labor market, organizational constraints, normative information). Cutoff scores are not necessarily criterion referenced, and different organizations may establish different cutoff scores on the same selection procedure based on their needs.
- Derived score** A score that results from a numerical transformation (e.g., conversion of raw scores to percentile ranks or standard scores) of the original selection procedure score.
- Differential item functioning** A statistical property of a test item in which different groups of test takers who have the same standing on the construct of measurement have different average item scores or, in some cases, different rates of endorsing various item options. Also known as DIF.
- Differential prediction** The case in which use of a common regression equation results in systematic nonzero errors of prediction for subgroups.
- Discriminant validity evidence** Evidence of a lack of relationship between measures intended to represent different constructs.
- Fairness** There are multiple perspectives on fairness. There is agreement that issues of equitable treatment, predictive bias, and scrutiny for possible bias when subgroup differences are observed are important concerns in personnel selection; there is not, however, agreement that the term “fairness” can be uniquely defined in terms of any of these issues.
- Generalized evidence of validity** Evidence of validity that generalizes to setting(s) other than the setting(s) in which the original validation evidence was documented. Generalized evidence of validity is accumulated through such strategies as transportability, synthetic validity/job component validity, and meta-analysis.
- Internal consistency reliability** An indicator of the reliability of a score derived from the statistical interrelationships of responses among item responses or scores on different parts of an assessment.
- Internal structure validity evidence** Demonstration of the degree to which psychometric and statistical relationships among items, scales, or other components within a selection procedure are consistent with the intended meaning of scores on the selection procedure.
- Inter-rater agreement** The consistency with which two or more judges rate the work or performance of examinees.
- Item** A statement, question, exercise, or task on a selection procedure for which the test taker is to select or construct a response, or perform a task.

**Item response theory (IRT)** A mathematical model of the relationship between performance on a test item and the test taker's standing on a scale of the construct of measurement, usually denoted as  $\theta$ . In the case of items scored 0/1 (incorrect/correct response) the model describes the relationship between  $\theta$  and the item mean score ( $P$ ) for test takers at level  $\theta$ , over the range of permissible values of  $\theta$ . In most applications, the mathematical function relating  $P$  to  $\theta$  is assumed to be a logistic function that closely resembles the cumulative normal distribution.

**Job component validity** See *Synthetic validity evidence*.

**Job description** A statement of the work behaviors and activities required or the worker requirements (e.g., knowledge, skills, abilities, and other personal characteristics).

**Job Knowledge** Information (often technical in nature) needed to perform the work required by the job.

**Job relatedness** The inference that scores on a selection procedure are relevant to performance or other behavior on the job; job relatedness may be demonstrated by appropriate criterion-related validity coefficients or by gathering evidence of the job relevance of the content of the selection instrument, or of the construct measured.

**KSAOs** Knowledge, skills, abilities, and other personal characteristics required in completing work in the context or environment in which an organization and individual may operate.

**Local evidence** Evidence (usually related to reliability or validity) collected in a single organization or at a specific location.

**Local study (local setting)** See *Local evidence*.

**Measurement bias** See *Bias*.

**Meta-analysis** A statistical method of research in which the results from several independent studies of comparable phenomena are combined to estimate a parameter or the degree of relationship between variables.

**Moderator variable** A variable that affects the strength, form, or direction of a predictor-criterion relationship.

**Multiple-hurdle model** The implementation of a selection process whereby two or more separate procedures must be passed sequentially.

**Normative** Pertaining to norm groups or the sample on which descriptive statistics (e.g., mean, standard deviation, etc.) or score interpretations (e.g., percentile, expectancy, etc.) are based.

**Norms** Statistics or tabular data (often raw and percentile scores) that summarize performance of a defined group on a selection procedure.

**Objective** Pertaining to scores obtained in a way that minimizes bias or error due to different observers or scorers.

**Operational setting** The specific organization, work context, applicants, and employees to which a selection procedure is applied.

**Personal characteristics** Traits or dispositions that describe individuals.

- Population** The universe of cases from which a sample is drawn and to which the sample results may be projected or generalized.
- Power** The probability that a statistical test will yield statistically significant results if an effect of specified magnitude exists in the population.
- Predictive bias** The systematic under- or overprediction of criterion performance for people belonging to groups differentiated by characteristics not relevant to criterion performance.
- Predictive validity evidence** Demonstration of the relationship between selection procedure scores and some future work behavior or work outcomes.
- Predictor** A measure used to predict criterion performance.
- Predictor-criterion relationship** The relationship between a predictor and external criteria (e.g., job performance, tenure) or other predictors and measures of the same construct.
- Professional judgment** Evaluations and decisions that are informed by and representative of the profession's commonly accepted empirical, methodological, and experiential knowledge base.
- Psychometric** Pertaining to the measurement of psychological characteristics such as aptitudes, personality traits, achievement, skill, and knowledge.
- Reliability** The degree to which scores for a group of assesseees are consistent over one or more potential sources of error (e.g. time, raters, items, conditions of measurement, etc.) in the application of a measurement procedure.
- Reliability estimate** An indicator that reflects the degree to which scores are free of measurement error.
- Response process** A component, usually hypothetical, of a cognitive account of some behavior, such as making an item response.
- Restriction of range or variability** Reduction in the observed score variance of a sample, compared to the variance of an entire population, as a consequence of constraints on the process of sampling.
- Sample** A selection of a specified number of entities called sampling units (test takers, items, etc.) from a large 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.
- Sampling bias** The extent to which a sampling process introduces systematic misrepresentation of the intended population.
- Score** A number describing the assessment of an individual; a generic term applied for convenience to such diverse kinds of measurements as tests, production counts, absence records, course grades, ratings or other selection procedures or criterion measures.

**Selection procedure** An assessment instrument used to inform a personnel decision such as hiring, promotion or placement.

**Selection procedure (test) user** The individual(s) or organization that selects, administers, and scores selection procedures (tests) and usually interprets scores that are obtained for a specified purpose in a defined organizational context.

**Shrinkage formula** An adjustment to the multiple correlation coefficient for the fact that the beta weights in a prediction equation cannot be expected to fit a second sample as well as the original.

**Skill** Level of proficiency on a specific task or group of tasks.

**Standardization** (a) In test construction, the development of scoring norms or protocols based on the test performance of a sample of individuals selected to be representative of the candidates who will take the test for some defined use; (b) in selection procedure administration, the uniform administration and scoring of a selection procedure in a manner that is the same for all candidates.

**Standard score** A derived score resulting in a distribution of scores for a specified population with specified values for the mean and standard deviation. The term is sometimes used to describe a distribution with a mean of 0.0 and a standard deviation of 1.0.

**Statistical power** See *Power*.

**Statistical significance** The finding that empirical data are inconsistent with a null hypothesis at some specified probability level.

**Subject matter experts** Individuals who have thorough knowledge of the work behaviors, activities, or responsibilities of job incumbents and the KSAOs needed for effective performance on the job.

**Synthetic validity evidence** Generalized evidence of validity based on previous demonstration of the validity of inferences from scores on the selection procedure or battery with respect to one or more domains of work (job components); also referred to as “job component validity evidence.”

**Systematic error** A consistent score component (often observed indirectly), not related to the intended construct of measurement.

**Test** A measure or procedure in which a sample of an examinee’s behavior in a specified domain is obtained, evaluated, and scored using a standardized process.

**Test development** Process through which a test or other predictor is planned, constructed, evaluated, and modified, including consideration of content, format, administration, scoring, item properties, scaling, and technical quality for its intended purpose.

**Trait** An enduring characteristic of a person that is common to a number of that person’s activities.

**Transportability** A strategy for generalizing evidence of validity in which demonstration of important similarities between different work set-

tings is used to infer that validation evidence for a selection procedure accumulated in one work setting generalizes to another work setting.

**Type I and Type II errors** Errors in hypothesis testing; Type I error involves concluding that a significant relationship exists when it does not; Type II error involves concluding that no significant relationship exists when it does.

**Validation** The process by which evidence of validity is gathered, analyzed, and summarized. (Note: laypersons often misinterpret the term as if it implied giving a stamp of approval; the result of the research might be zero validity.)

**Validity** The degree to which accumulated evidence and theory support specific interpretations of scores from a selection procedure entailed by the proposed uses of that selection procedure.

**Validity argument** An explicit scientific rationale for the conclusion that accumulated evidence and theory support the proposed interpretation(s) of selection procedure scores entailed by the proposed uses.

**Validity coefficient** A measured coefficient reflecting the relationship between a selection procedure and a criterion that provides evidence about the validity of the selection variable.

**Validity evidence** Any research or theoretical evidence that pertains to the interpretation of predictor scores, or the rationale for the relevance of the interpretations, to the proposed use.

**Validity generalization** Justification for the use of a selection procedure or battery in a new setting without conducting a local validation research study. See *Generalized evidence of validity*.

**Variable** A quantity that may take on any one of a specified set of values