Psychometric/Assessment Terms

Psychometric and Measurement Terms
Fall, 2005

Psychometric/Measurement Concepts

Accuracy: (1) The extent to which observed values approximate the “true” state of nature (Foster & Cone, 1995; Johnston & Pennypacker, 1993; Silva, 1993). (2) Correctness. Precision. Exactness (James & James, 1990). (3) The degree to which obtained data are correct measures of targeted phenomena. (4) The degree to which the measurement represents the objectively defined topographic characteristics of the measured event (Cone, 1988). Used most often but not exclusively with behavioral observation.

Agreement: 1. The degree of correspondence between the output of two or more assessment instruments. 2. The degree of overlap between two more independent records (Silva, 1993).

Clinical applicability (of an assessment instrument): The clinical utility of an assessment instrument across clinical populations, settings, ages, and other sources of individual differences.

Criterion Contamination—Occurs when elements of an assessment instrument are highly correlated elements of the instrument used for its validation (e.g., items from the two instruments have identical wordings). Measures of covariance then partially represent test-retest reliability.

Homogeneity: The degree to which elements of a class are the same or similar. It is often estimated through internal consistency analyses.

Internal Consistency: The degree of consistency of the items or elements within an assessment instrument. Can be reflected by split-half reliability, Kuder-Richardson 20 formula, coefficient alpha, item-total, (Kazdin, 1992).

Interobserver, Interrater, Interscorer Agreement: (1) The extent of agreement between scores (or ratings, diagnoses, behavior rates) obtained from different assessors. Can be indicated by percent agreement, kappa, or correlation. (2) The extent to which different assessors agree on the scores they provide when assessing (e.g., rating, coding, classifying) a subjects performance (Kazdin, 1992).

Measure: (1) A number that represents the attribute being measured. The score obtained from an assessment instrument (e.g., blood pressure reading, MMPI scale score, observed rate of behavior. (2) A system of units with which attributes can be described (e.g., ounces, rate, degrees, lbs) (James & James, 1992).

Measurement: (1) The assignment of a numerical value to variable dimension. (2) the process of assigning numbers or other symbols to the things in such a way that relations of the numbers or symbols reflect relations of the attribute being measured (Sarle, e-mail, 1/7/95).
Psychometric/Assessment Terms

**Measurement Error**: The part of an obtained measure that does not reflect true variance in the targeted variable.

**Precision, of a measure**: The degree to which a class of measurements approximate the “true” value and estimated through coefficients of internal consistency, temporal stability, and dispersion (Marriott, 1990).

**Reliability**: 1. The part of test result that is due to permanent, systematic effects and therefore persists from sample to sample (Marriott, 1990). 2. the stability of measurement under constant conditions (Silva, 1993). Precision of a measure; replicability.

**Sensitivity to Change**: The degree to which measures from an instrument reflect changes across time in the targeted construct.

**Shared Method Variance (Common Method Variance)**--The similarity in the procedures used to acquire data; can contribute to the magnitude of correlation between data from different assessment instruments (e.g., both instruments are self-report; both instruments use the same informant; shared items);

**Validation (Validity Assessment)**: (1) The process of establishing the validity of data from an assessment instrument (e.g., Content validation is the process of establishing the content validity. . .) (2) The process of evaluating the validity of inferences based on the scores from an assessment instrument (Cronbach, 1971).

**Validity**: An integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on the data acquired from an assessment instrument (Messick, 1993). (2) The scientific utility of an assessment instrument; how well it measures what it purports to measure (Nunnally & Bernstein, 1994). (3) The meaning of scores from an assessment instrument (Foster & Cone, 1995; Cronbach & Meehl, 1955). Closest in meaning to “construct validity.”

**Validity, Concurrent**: (1) Any index of validity obtained when multiple measures are administered on the same assessment occasion. (2) The correlation of a measure or criterion at the same point in time (Kazdin, 1992).

**Validity, Construct**: (1) Comprises the evidence and rationales supporting the trustworthiness of assessment instrument data interpretation in terms of explanatory concepts that account for both the obtained data and relations with other variables (Messick, 1993). (2) the degree of validity of inferences about unobserved variables (constructs) based on observed indicators (Pedhazur & Schmelkin, 1991); (3) the extent to which the measure assesses the domain, trait, or characteristic of interest; refers broadly to the evidence bearing on the measure and encompasses all types of validity (Kazdin, 1992)

**Validity, Convergent**: (1) The degree to which the data from the assessment instrument are coherently related to other measures of the same construct as well as to other variables that it is expected, on theoretical grounds, to be related to (Messick, 1993). (2) The extent of covariance between scores from two assessment instruments that measure the same or related
constructs. The correlation between the measures is expected based on the overlap or relation between the constructs. (Campbell, & Fiske, 1959)

Validity, Content: The degree to which elements of an assessment instrument are relevant to and representative of the targeted construct, for a particular assessment purpose (Haynes, Richard, & Kubany, 1995)

Validity, Criterion-referenced: (Criterion-related validity; Criterion validity) (1) The degree to which measures from an assessment instrument correlate with scores from previously validated instruments that measure the phenomena of interest or with nontest criterion of practical value. (2) Correlation of a measure with some other (validated) criterion. This can encompass concurrent or predictive validity (Kazdin, 1992).

Validity, Discriminant (divergent): (1) The degree to which data from an assessment instrument are not related unduly to other exemplars of other constructs (Messick, 1993). (2) The correlation between measures that are expected not to relate to each other or to assess dissimilar or unrelated constructs (Kazdin, 1992). Discriminant validity of an instrument is suggested by small or no significant correlation with data from instruments that tap dissimilar constructs. (Campbell & Fiske, 1959). Most useful when applied to constructs that should not, but could, account for variance in the primary measure of interest.

Validity, Discriminative: The degree to which measures from an assessment instrument can differentiate individuals in groups, formed from independent criteria, known to vary on the measured construct. For example, the ability of a score from a marital inventory to differentiate individuals who are and who are not seeking marital counseling.

Validity, Divergent: see “Validity, Discriminant”.

Validity, Ecological: (see Barkley article; in analog ass folder)

Validity, Face: A component of content validity. It refers to the degree that respondents or users judge that the items of an assessment instrument are appropriate to the targeted construct and assessment objectives (Allen & Yen, 1979; Anastasi, 1988; Nevo, 1985). It is commonly thought to measure the “acceptability” of the assessment instrument to users and administrators (Haynes, Richard & Kubany, 1995).

Validity, Incremental: The degree to which data from an assessment instrument/process increase the validity of judgments beyond that associated with assessment instruments/processes currently in use or beyond that associated with alternative assessment instruments/processes. (Sechrest, 1963; Haynes and Lench, 2003).

Validity, Postdictive: The degree to which scores from an assessment instrument correlate with scores from another validated assessment administered at a previous point in time, or the degree to which scores predict historical events.

Validity, Predictive: (1) The degree to which scores from an assessment instrument correlate with scores from another, validated assessment administered at a later point in time. (2) The degree to which scores from an instrument estimate an external criterion (Nunnally &
Psychometric/Assessment Terms

Bernstein, l994; note that the time frame of prediction is not important in this definition; the “criterion” is the determining factor).

Validity: Treatment: The degree to which data from an assessment instrument(s) contribute to enhanced treatment outcome. (Hayes, Nelson, & Jarret, (1987); Silva, l993). Also termed “treatment utility”.

Concepts of Causation of Behavior Problems (which are relevant to discussions of functional relations). See Causality Glossary for more and refined definitions.

Behavior disorder/Behavior problem: Generic terms of convenience used to refer to any target of clinical intervention. The use of these terms does not imply that the disorder/problem is an empirically validated construct or that it is composed of a reliably covarying set of behaviors, such as is implied in DSM-III-R. It also does not imply that there is an "underlying" cause to a group of "symptoms". Thus "behavior disorders" or "behavior problems" may be used, interchangeably, at a variety of levels to refer to self-stimulatory behaviors, self-reported depressed affect, amount of exercise, negative self-statements, excessive caloric intake.

Causal variable: A variable that controls a proportion of the variance in another variable; causal variables precede and are correlated with their effects, have a logical connection with their effects, and the association between the causal variable and the effect cannot be wholly attributed to a common effect of another variable.

Domain (of a causal relation): Conditions under which a causal relation, or a particular form or strength of the causal relation, is operational. Domains may involve developmental stages, environmental settings and contexts, physiological states, parameter values, time, persons, or the presence or absence of particular variables.

Functional relation: A functional relation exists when two or more variables have shared variance: Some dimension (e.g., rate, magnitude, length, age) of one variable is associated with some dimension of another. In an alternative language, variables are functionally related when they demonstrate a mathematical relation (Haynes, 1992). A mathematically describable relation between two or more variables.

Level of Inference (specificity): The number of elements or components subsumed by the variable label (Haynes et al., 1993, 1994). Behavior problems, goals and causal variables at higher or lower levels (sometimes termed “molar” or “molecular”; “higher order” or “lower order”). A higher level behavior problem is illustrated by “depression,” that can refer to multiple lower-level phenomena such as motor slowness, negative affect, insomnia, and eating disturbances.

Mediating variable: A variable that accounts for, or explains, the relation between two other variables; similar to a “causal mechanisms” (Baron & Kenny, 1986; Shadish, 1996)
**Psychometric/Assessment Terms**

**Moderator Variable:** A variable that can influence the strength and or direction of the relation between two or more other variables (Baron & Kenny, 1986). A variable whose value covaries with (affects) the magnitude of a causal relation (Shadish, 1996).

**Time-course:** (a) The values of a variable dimension as a function of time; (b) the temporally related dimensions of a variable, such as cyclicity, latency, duration, and rate. The time-courses of variables are frequently presented in graphical form with time on the horizontal axis and the value of the variable on the vertical axis. (Haynes et al., 1995).

**Clinical Judgment, Utility, Assessment Concepts**

**Clinical Significance:** The extent to which an obtained effect (e.g., intervention effect, shared variance between variables) is important, has practical value, or can guide clinical judgments.

**Clinical utility** (of an assessment instrument): The degree to which the results of an assessment instrument enhance the validity of clinical judgments. The clinical utility of an assessment instrument can vary across its applications (e.g., diagnosis, brief screening).

**Conditional Probability:** If A and B are events, the conditional probability of A given B is the probability of A, assuming B holds. If P(B) ≠ 0, then the conditional probability of P(A/B) of A given B is P(A and B)/P(B) (James and James, 1992; see Schlundt, 1985, for example of clinical application).

**Cost-effectiveness, of an assessment instrument:** The cost (e.g., time, financial) of deriving information with an assessment instrument relative to the contribution of that information to a clinical judgment.

**Effect size:** (1) A coefficient of the statistical relation between two variables. It is usually expressed as some form of “proportion of variance accounted for”. (2) In treatment outcome research it is a way of expressing the difference between two groups in a metric that can be applied across studies (e.g., difference between means divided by pooled standard deviation) (Kazdin, 1992). (3) the mean difference between the control group and the treatment group [(mean1 - mean2)/SD of control group] or [(mean1 - mean2)/pooled SD] and then adjusted by instrument reliability coefficient.

**Magnitude of effect:** A measure of the strength of relation between variables, or the magnitude of change across time, or the magnitude of difference between groups. Indices of the magnitude of effect include “r,” “R^2” (see Effect size).

**Negative Predictive Power:** (1) true negative rate/true negative + false negative rate; the proportion of individuals indicated as not having a disorder/behavior who truly do not; Sometimes confused with “specificity”.

**Positive Predictive Power:** (1) indicates the “hit rate” for an assessment instrument score; true positive rate/true positive + false positive rate; e.g., the proportion of individuals identified by an instrument who truly have the disorder/behavior Sometimes confused with “sensitivity”. 

Power (of an assessment instrument): (1) The predictive accuracy of an assessment instrument. Usually estimated by the proportion of persons it accurately identifies with (sensitivity) and without (specificity) an attribute (such as a diagnosis). (2) The probability of detecting a difference between groups when a difference truly exists (Kazdin, 1992).

Sensitivity: The probability that a person with a particular attribute will manifest a particular behavior. The proportion of positive cases so identified by an assessment instrument. See also Positive Predictive Power.

Specificity, of an assessment instrument: The probability that a person without a particular attribute will be so identified by a particular assessment instrument. The proportion of negative cases so identified by an assessment instrument. See also “Negative Predictive Power”

Specificity, of a measure or variable: The degree of molarity or precision of a measure and can refer to (a) The diversity and number of elements subsumed by a measure, (b) The degree to which the dimensions or parameters of a variable are specified, (c) The degree to which situational and temporal conditions relevant for the target variable are specified; and (d) The level of specificity of clinical judgments based on obtained measures. (Haynes, Nelson & Blaine, 1999).

Additional TermnAssessment Strategies

Functional assessment: (1) An identification of the antecedent and consequent events that are temporally contiguous to the target response, and that occasion and maintain it (Sisson & Taylor, 1993); similar to definition for functional analysis; (2) An assessment of functional capabilities; this definition is often used in rehabilitation psychology and neuropsychology;

Idiographic Assessment: Assessment pertaining to an individual or individual case. Refers to assessment procedures that are not standardized, and observed relations and results that are not necessarily generalizable, across persons or groups. Sometimes called idiothetic (e.g., Tallent, 1992) and incorrectly labelled ideographic.

Informants: Persons furnishing information about a participant/patient; typically those with frequent contact, such as parents, teachers, spouses.

Measurement/Assessment Strategy: (1) The procedures used to acquire data. (2) The assessment instruments used to acquire data and the methods in which they are applied; These methods may include time, behavior and situation sampling parameters.

Multiinformant Assessment: Assessment that includes more than one informant, such as the client, parents, teachers, spouse, and staff members.

Multimethod Assessment: Assessment that includes more than one method of gathering information, such as self-monitoring, psychophysiological assessment, and interviews.

Multimodal Assessment: Assessment that targets more than one response mode, such as
Psychometric/Assessment Terms

thoughts, behavior, physiological responses, and emotions.

**Multisource Assessment:** Assessment that includes multiple sources of information, such as multiple methods, instruments, occasions, situations, and informants.

**Nomothetic, Assessment Strategy:** An assessment strategy in which judgments are based on the comparison of measures from the target person with data on the same instrument gathered from other persons, such as the use of normative or comparison groups.

**Response Mode (Response System):** The form, type, or method of behavior. Response modes are organizational categories or a taxonomy of behavior. Response modes and can include motor, verbal, cognitive, physiological and combinations of these, such as emotional.

**Time-Series Assessment:** Includes a diverse set of assessment strategies to describe and analyze the time-courses and interrelations of multiple variables. With time-series assessment, behavior problems or hypothesized causal variables are measured frequently (e.g., 30 or more measurements) across time. Measurement occurs at a sufficient rate, and at sufficiently short intervals, to detect serial correlation in the time series.