

Assessment and Measurement Terminology
MUE 6747 – Assessing Music Learning – Dr. Brophy

Terms

1. *Distribution*—the values obtained from the measurement process for a specific population
2. *Levels of measurement*
 - a. Nominal—pertaining to the assignment of numbers as names or labels in order to categorize or sort the data; i.e., gender classification as male = 1 and female = 2
 - b. Ordinal—having to do with the assigning of numbers to indicate rank orders without implication for the amount of difference between the ranks (i.e., chair assignments in orchestra or band)
 - c. Interval—the assignment of a constant unit of measurement that serves as an interval (such as a point), so that differences between scores can be determined; such scales have an arbitrary zero point. This type of system is most common in standardized music testing, and the numbers obtained are subject to statistical analysis
 - d. Ratio—a measure with an absolute zero point that represents the genuine absence of the property in question; one can make “twice as much” and “half as much” statements (etc.) with data measured from these types of instruments
3. *Central tendency*—mathematical descriptions of the point around which a set of values in any population distribution tend to cluster
 - a. Mean—the arithmetic average of a set of values
 - b. Median—the midpoint of a distribution
 - c. Mode—the most commonly occurring value in a distribution
 - d. Skew—a description of a distribution’s symmetry; if a distribution’s mean is higher than its median, the distribution is *positively skewed*; if the mean happens to be below the median, then the distribution is *negatively skewed*
4. *Variability*—having to do with how a distribution’s values vary among themselves
5. *Variation*—a measure of how a distribution is spread (or scattered); it is the mean of the squared deviations from the distribution’s mean (s^2)
 - a. Deviation—the arithmetic difference between a particular value in the distribution and its mean
6. *Standard Deviation*—a linear measure of the distribution of values, the standard deviation (s) is the square root of the variance (s^2); the standard deviation is expressed in terms of the distribution’s original unit of measurement
7. *Standard Error of Measurement*—an estimate of the variability in the expected distribution of a given individual’s score if that individual’s particular knowledge, skill or attitude were to be measured repeatedly many times; these scores would vary around a theoretical “true score”
8. *Range*—the difference between the highest and lowest value in a distribution
9. *Reliability*—the consistency with which a measuring instrument measures, expressed as the proportion of variance in a set of values that is due to genuine variability in what is being measured; ranges between 00.0 (perfectly unreliable) and 1.00 (perfectly reliable). The closer to 1.00 the reliability estimation, the more reliable the test.

- a. *Correlation coefficient*—a number that expresses the size and direction of a relationship between two sets of scores (this is a linear relationship); expresses parallel forms reliability, test-retest reliability, and split-halves reliability
 - i. Parallel forms reliability—an estimate of the reliability between two forms of the same test
 - ii. Test-retest reliability—correlation of scores from two administrations of the same test
 - iii. Split-halves reliability—most often used for one-time teacher made tests, the test is divided into two sections and the scores for each section are correlated
 - iv. Internal consistency methods—the relationship between individual items (interitem consistency); coefficient α for variable point items, and KR₂₀ for dichotomous items
10. *Validity*—the degree to which a measure measures what it is supposed to measure; determined qualitatively
- a. *Content validity*—has to do with how the measure covers the content it is supposed to sample
 - b. *Predictive validity*—having to do with how well the measure predicts a particular outcome (future success)
 - c. *Concurrent validity*—the use of correlation between scores on a new test and an older, validated test as a measure of the validity of the newer test
 - d. *Construct validity*—having to do with how well the measure yields results that are predicted or presupposed by a theory (a construct)
11. *Norms*—“representative scores” that facilitate comparison among individuals, established by administering the test to a large representative population and equating the raw scores with one or more forms of equivalent scores
- a. *Standard score*—based on the number of standard deviations by which a score deviates from the mean of a given distribution; the basic form is the *z-score*; another form is the *T-score*
 - b. *Age score*—age equivalents that are most common for particular raw scores
 - c. *Percentile rank*—the percent of a distribution of scores that a particular score meets or exceeds
12. *Criterion-referenced reliability measures*—used to determine the consistency of “mastery-nonmastery” classifications; the most common is Subkoviak’s classification consistency statistic, used for classification consistency between two forms of a criterion-referenced test, and coefficient kappa is used to correct this for chance classifications; others include Brennan’s procedures and Livingston’s reliability