Assessment

Comps Study Guide #6

Define the following terms:
appraisal,
measure,
test,
objective test,
subjective test,
projective test,
free choice test,
and forced choice items.

Appraisal - a variety of assessment tools including tests and surveys used to evaluate traits and behaviors.
Measure - a score assigned to a person's traits or behavior.
Test - a systematic method of measuring or evaluating.
Objective test items - based on a universal standard such as multiple choice - require little or no judgment in scoring.
Subjective test items - items such as essay questions - scoring of these items requires judgment and may reflect the scorer's bias.
Projective tests - unstructured tests that may reveal basic personality, concealed feelings, and internal conflicts.
Free choice test - short answer questions that elicit subjective information.
Forced choice items - items such as true/false questions for which the test taker must recall information

Define the following terms: spiral test, cyclical test, test battery, horizontal test, vertical test and Q-Sort.

Spiral test - starts with easier questions and progresses to the harder ones.
Cyclical test - test has multiple sections and the questions in each section progress in difficulty.
Test battery - a collection of tests given to the same group of people and scored against the same standard.
Horizontal test - a test procedure that covers material from different subjects.
Vertical tests - tests on the same subject given at different levels or ages.
Q-Sort - a tool for measuring self-esteem by choosing statement-bearing words that are "most like me" or "least like me."

Define the following terms:
halo effect,
difficulty index,
dichotomous items,
normative,
ipsative format,
power test,
and speed test.

Halo effect - a favorable evaluation of a personality based on the perception of a single trait.
Difficulty index - in testing the percentage of test takers who respond correctly to an item.
Dichotomous items - questions such as true/false that give the test taker opposing choices.
Normative item format - unlinked items on a test
Normative test - a person's test results can be compared to the scores of others - a percentile rank can be created.
Ipsative format - allows a person to compare two or more examples of his/her own performance - does not allow for comparison with others.
Power test - untimed test- tests mastery level.
Speed tests - timed test - difficulty is more in how quickly questions can be answered than in the content.

Define the following terms:
percentile,
stanine,
z score,
T-score,
standardized score and
standardized score.

Percentile - on a scale of 100 the number that shows the percent of a lata distribution equal to or below it
Stanine, or Standard Nine - a way of scaling test scores - nine divisions, five of them in the middle with a standard deviation of 2.
The lowest z-scores comprise the first group and the highest scores the last.
T-score - a method for determining a standardized score - subtract the mean from an individual score then divide by the standard deviation.
Standardized score - a score within a normal distribution with a mean of 50 and a standard deviation of 10.
Standardized score - same as z-score

Define the following terms:
mean,
median,
mode,
Skew,
range,
standard deviation,
variance,
and bell curve.

Mean - the average score from a group of tests.
Median - the middle score from a group of tests.
Mode - the score that occurs most frequently in a set of scores.
Skew - the amount a score deviates from the norm.
Range - the lowest score subtracted from the highest.
Standard deviation - a measure of statistical dispersion - in testing, how widely spread the scores are from the mean.
Variance - the square of the standard deviation.
Bell curve - a graphic illustration of the normal distribution of a data set

Describe the following types of reliability:
stability,
equivalence,
and internal consistency.

Test reliability is a measurement of the consistency with which a test yields similar results in repeated uses.
Stability - also called test-retest reliability - giving the same test to the same group twice with no more than two weeks between so that the two sets of results can be correlated without intervening experiences affecting the outcome.
Equivalence - the correlation of the results of using different tests covering the same content with the same group of test takers. Time between the tests and the formats used can affect the outcome.
Internal consistency - measures the consistency of results from items of a test - do the responses from similar and opposing questions yield consistent information.

Define the following terms:
correlation coefficient,
reliability,
intrusive measurement
and obtrusive measurement.

Correlation coefficient - a measurement of the linear relationship between two variables.

Reliability - the consistency with which a test yields similar results, measured by the use of a correlation coefficient. Reliable does not equate to validity.

Intrusive measurement - questionnaires, interviews, and other situations in which a person is aware s/he is being observed - that awareness can affect the results of the observation. Measurement can also be called reactive.

Obtrusive measurement - (nonreactive), subject is unaware of observation or investigation, such as when records are reviewed or subject is observed through a one-way window.

Explain what validity is
and define
face and content validity.

The validity of a test is the degree to which it measures what it is designed to measure. The content of the test must accurately measure the skills or information learned. The validity of a test is specific to a situation, including why and to whom it is administered. A testing instrument that is valid for one situation or population may not be valid for another. A valid test must always be reliable.

Face validity is obvious validity, for example the questions on a math test will deal with math.

Content validity, which can also be called rational or logical validity, is the reflection of the subject matter in the content of the test, for example a math test will contain material covered in the specific math course.

Define coefficient of determination
and the
standard error of measurement.

Coefficient of determination - the square of the correlation coefficient which shows the common variation between the two variables - in testing, the amount of common variance between repeated tests.

Standard error of measurement (SEM) - a statistical range that will include a test taker's score - calculated by the multiplication of the test's standard deviation by the square root of it, then the subtraction of the reliability coefficient.

List the circumstances when testing is useful and give examples of the purpose for administering tests to clients.

Tests are used for many purposes in many different situations.

- In educational institutions they measure academic achievement
- In job placement, can help guide people to careers for which they are best suited.

They can also be used to predict future performance and success in both education and work.

Tests are also used to ensure the qualifications of persons who apply for licenses or certification in a field if employment

Counselors use various tests to evaluate clients and to help the clients learn about themselves. Among the reasons a counselor nay administer a test to a client are:

- to determine if the client's needs are within the scope of the counselor's practice,
- to help the client understand himself or herself,
- to help the counselor better understand he client,
- to determine which methods and techniques are most appropriate for a particular client,
- to aid the client in decision making,
- to identify interests, and to evaluate the counseling.

Explain predictive, concurrent and construct validity.

Predictive validity: which can also be called empirical validity, is the capability of a testing instrument to predict future behavior, for example the ability of the Graduate Record Exam to predict a person's grade point average.

Concurrent validity is the immediate comparison of test results with the results from other sources that measure the same factors in the same short time span.

Construct validity is the extent to which a testing instrument measures an abstract psychological trait such as anxiety.

List the steps in interpreting test scores with a client.

(1) The counselor should be trained in test theory, and before administering the test, should study the technical manual for the test
(2) The counselor should understand the scores, profiles, and implications of the test
(3) Using non-technical language, the counselor should explain the test to the client, including the reason for the test and what it measures.
(4) When reviewing the scores with the client, the counselor should explain percentiles and other technical terms.
(5) The results of the test should be presented to the client in an organized manner and in layman's terms. The interrelationship of the multiple tests should be explained if more than one was used.
(6) The counselor should help the client to integrate the results of the test with other factors and encourage the client to express reactions and emotions.
(7) The counselor should assure the client that test scores are just tools to help with decision making and not infallible limits placed on him or her.
(8) The interpretation session should not be rushed; time should be available for the client to ask questions and discuss the results.

List the major types of tests and inventories with examples.

Intelligence exams are used to measure a person's mental ability. Stanford-Binet Intelligence Scales, Wechsler Adult Intelligence Scale (WAIS-III), and Kaufman Assessment Battery for Children are examples of this type.

Achievement tests measure learning and are often given in schools at particular grade levels or as "end of year" tests. They can also be used as diagnostic tools. Some examples are: the California Achievement Test and the General Education Development {GED).

Aptitude tests, which can also be called ability tests, are used to measure a person's ability to master skills or acquire knowledge. Differential Aptitude Tests {DAT] and Career Ability Placement Survey (CAPS) are examples.

Personality tests are used to determine a person's personality traits and may be projective, inventories or specialized. Examples are: projective - Rorschach, and Thematic Apperception Test (TAT);

Personality inventory - Minnesota Multiphasic Personality Inventory (MMPI-2) and Myers Briggs Type Indicator; s

Specialized - Tennessee Self-concept Scale and Luria-Nebraska Neuropsychological Battery.

Interest inventories are used to determine a person's likes and dislikes. Strong Interest Inventory, Career Assessment Inventory, and O*Net Interest Profiler are examples.

Define the terms:
regression to the mean,
rating scale,
sociometry,
and psychometric.

Regression. to the mean is a statistical concept where earning a very low score or a very high score on a pretest means the individual will probably score close to the mean on the posttest. The error is due to chance, personal and environment factors that will be different on the posttest.

Rating scale is a cart used to indicate the degree to which an attribute or characteristic exists.

Sociometry was coined by Joseph Levy Moreno and is a method of tracking the relationship of individuals within a group. A **sociogram** is a map or diagram showing the structure of the group or relationships of the members.

Psychometric refers to any form of mental testing.

Explain ethical issues in testing.

Confidentiality of test results is of primary importance since some test results can label or stereotype the test taker, and in some instances, may be an invasion of privacy. Confidentiality is especially important if test records are placed on a computer. The counselor must maintain security measures, both physical and on the computer, to ensure that the records are not accessible to unauthorized persons.

Another issue concerns the tests themselves since most were developed to test white, middle-class males and may not accurately test females, non-white males, or persons from minority cultures. The counselor should make every effort to see that the tests he or she uses are as unbiased as possible. Some issues of validity, interpretation, and confidentiality concern computer administered tests, which can include self-testing.

Explain why WAIS-III is better for adults.

David Wechsler wanted to create a test that did not rely solely on verbal skills. The Wechsler Adult Intelligence Scale, WAIS-Ill, is one of the IQ tests developed by David Wechsler and administered on an individual basis. The WAIS-III provides a verbal IQ, performance IQ, and a full-scale IQ. There are a total of 7 verbal plus 7 performance scales.
Wechsler also developed the WPPSI-R, Wechsler Preschool and Primary Scale of intelligence, for children ages 3 years to 7 years/3 months.

For children7 years to 16 years/11months, he developed the Wechsler Intelligence Scale for Children Revised, WISC-III.

Describe the contribution of Frances Galton, J.P. Gilford, Alfred Binet, and M. Terman.

Sr. Francis Galton is recognized as the leading pioneer in the study of individual differences. He concluded that intelligence was primarily genetic and had a normal distribution, similar to height and weight.

J.P. Gilford used factor analysis and isolated 120 factors that added up to intelligence. He also defined convergent and divergent thinking. Convergent thinking is when different thoughts and ideas are combined into a single concept. Divergent thinking is the ability to create a novel idea.

Alfred Benet, along with Theodore Simon, is credited with creating the first intelligence test The year was 1905 and the test consisted of 30 items of increasing difficulty administered to discriminate normal from retarded Parisian children.

It was adapted for America by **M. Terman** of Stanford University and became the Stanford-Binet IQ test. The original Stanford-Binet produced the intelligence quotient or ratio IQ (mental age divided by chronological age times 100: MA/CA x 100). Today the IQ formula is SAS, standard age score, with a mean of 100 and a standard deviation of 16.

Identify the following individuals:
John Ertl,
Raymond Cattell,
Arthur Jensen,
and Robert Williams.

John Ertl advanced the theory that the rate at which a person processes information is an indication of his or her level of intelligence - the faster the processing, the more intelligent the person. He also invented an intelligence-testing machine that uses an electrode helmet in conjunction with a computer and an EEG.

Raymond Cattell developed the theories of fluid and crystallized intelligence. Fluid intelligence is inborn, deals with abstract reasoning, is unrelated to experience, and decreases with age. Crystallized intelligence develops from acquired knowledge and skills. He is also the creator of the *16 Personality Factor Questionnaire.*

Arthur Jensen applied the theory that intelligence is genetic to adopted children, expecting them to have IQ scores closer to their biological parents than to their adoptive parents. He believed that 80% of intelligence is inherited and only 20% is environmental.

Robert Williams, an African-American psychologist, created the *Black intelligence Test of Cultural Homogeneity (BITCH*) as proof that African Americans can excel on intelligence tests when the cultural bias is toward their own experience rather than toward White culture.