Source: Cohen & Swerdlik (2018), Kaplan & Saccuzzo (2018), Groth & Wright (2016), Psych Pearls

Hellow, future RPm!

Doubt is definitely going to eat you this review season. I just want you to know that it is very VALID to feel that way. But always remember to go back to the reason why you are doing this.

May this reviewer help you pass the boards like it did to me and to many people that I know <3

We will be remembered ✨

Psychometric Properties and Principles (39)

Psychometric Properties essential in Constructing, Selecting, Interpreting tests

Psychological Testing - process of measuring psychology-related variables by means of devices or procedures designed to obtain a sample of behavior
- numerical in nature
- individual or by group
- administrators can be interchangeable without affecting the evaluation
- requires technician-like skills in terms of administration and scoring
- yield a test score or series of test score
- minutes to few hours

Psychological Assessment - gathering and integration of psychology-related data for the purpose of making psychological evaluation
- answers referral question thru the use of different tools of evaluation
- individual
- assessor is the key to the process of selecting tests and/or other tools of evaluation
- requires an educated selection of tools of evaluation, skill in evaluation, and thoughtful organization and integration of data
- entails logical problem-solving that brings to bear many sources of data assigned to answer the referral question
- *Educational:* evaluate abilities and skills relevant in school context
- *Retrospective:* draw conclusions about psychological aspects of a person as they existed at some point in time prior to the assessment
- *Remote:* subject is not in physical proximity to the person conducting the evaluation
- *Ecological Momentary:* "in the moment" evaluation of specific problems and related cognitive and behavioral variables at the very time and place that they occur
- *Collaborative:* the assessor and assesee may work as "partners" from initial contact through final feedback
- *Therapeutic:* therapeutic self-discovery and new understanding are encouraged
- *Dynamic:* describe interactive approach to psychological assessment that usually follows the model: evaluation > intervention of some sort > evaluation

- **Psychological Test –** device or procedure designed to measure variables related to psychology
  - *Content:* subject matter
  - *Format:* form, plan, structure, arrangement, layout
  - *Item:* a specific stimulus to which a person responds overtly and this response is being scored or evaluated
  - *Administration Procedures:* one-to-one basis or group administration
  - *Score:* code or summary of statement, usually but not necessarily numerical in nature, but reflects an evaluation of performance on a test
  - *Scoring:* the process of assigning scores to performances
  - *Cut-Score:* reference point derived by judgement and used to divide a set of data into two or more classification
  - *Psychometric Soundness:* technical quality
  - *Psychometrics:* science of psychological measurement
  - *Psychometrist or Psychometrician:* refer to professional who uses, analyzes, and interprets psychological data

Source: Cohen & Swerdlik (2018), Kaplan & Saccuzzo (2018), Groth & Wright (2016), Psych Pearls

Ability or Maximal Performance Test – assess what a person can do
1. **Achievement Test** – measurement of the previous learning
- used to measure general knowledge in a specific period of time
- used to assess mastery
- rely mostly on *content validity*
- fact-based or conceptual
2. **Aptitude** – refers to the potential for learning or acquiring a specific skill
- tends to focus on informal learning
- rely mostly on *predictive validity*
3. **Intelligence** – refers to a person's general potential to solve problems, adapt to changing environments, abstract thinking, and profit from experience

Human Ability – considerable overlap of achievement, aptitude, and intelligence test

Typical Performance Test – measure usual or habitual thoughts, feelings, and behavior
- indicate how test takers think and act on a daily basis
- use interval scales
- no right and wrong answers

Personality Test – measures individual dispositions and preferences
- designed to identify characteristic
- measured ideographically or nomothetically
1. **Structured Personality tests** – provide statement, usually self-report, and require the subject to choose between two or more alternative responses
2. **Projective Personality Tests** – unstructured, and the stimulus or response are ambiguous
3. **Attitude Test** – elicit personal beliefs and opinions
4. **Interest Inventories** – measures likes and dislikes as well as one's personality orientation towards the world of work

Other Tests:
1. **Speed Tests** – the interest is the number of times a test taker can answer correctly in a specific period
2. **Power Tests** – reflects the level of difficulty of items the test takers answer correctly
3. **Values Inventory**
4. **Trade Test**
5. **Neuropsychological Test**
6. **Norm-Referenced test**
7. **Criterion-Referenced Tests**

- **Interview** – method of gathering information through direct communication involving reciprocal exchange

Standardized/Structured – questions are prepared

Non-standardized/Unstructured – pursue relevant ideas in depth

Semi-Standardized/Focused – may probe further on specific number of questions

Non-Directive – subject is allowed to express his feelings without fear of disapproval

  - *Mental Status Examination:* determines the mental status of the patient
  - *Intake Interview:* determine why the client came for assessment; chance to inform the client about the policies, fees, and process involved
  - *Social Case:* biographical sketch of the client
  - *Employment Interview:* determine whether the candidate is suitable for hiring
  - *Panel Interview (Board Interview):* more than one interviewer participates in the assessment
  - *Motivational Interview:* used by counselors and clinicians to gather information about some problematic behavior, while simultaneously attempting to address it therapeutically
- **Portfolio** – samples of one's ability and accomplishment
- **Case History Data** – refers to records, transcripts, and other accounts in written, pictorial, or other form that preserve archival information, official and informal accounts, and other data and items relevant to an assessee
  - *Case study:* a report or illustrative account concerning a person or an event that was compiled on the basis of case history data
  - *Groupthink:* result of the varied forces that drive decision-makers to reach a consensus
- **Behavioral Observation** – monitoring of actions of others or oneself by visual or electronic means while recording quantitative and/or qualitative information regarding those actions
  - *Naturalistic Observation:* observe humans in natural setting
  - *SORC Model:* Stimulus, Organismic Valuables, Actual Response, Consequence
- **Role Play** – defined as acting an improvised or partially improvised part in a stimulated situation
  - *Role Play Test:* assesses are directed to act as if they are in a particular situation
- Other tools include computer, physiological devices (biofeedback devices)

Psychological Assessment Process

1. Determining the Referral Question

2. Acquiring Knowledge relating to the content of the problem

Source: Cohen & Swerdlik (2018), Kaplan & Saccuzzo (2018), Groth & Wright (2016), Psych Pearls

3. Data collection

4. Data Interpretation

- **Hit Rate –** accurately predicts success or failure
- **Profile –** narrative description, graph, table. Or other representations of the extent to which a person has demonstrated certain targeted characteristics as a result of the administration or application of tools of assessment
- **Actuarial Assessment –** an approach to evaluation characterized by the application of empirically demonstrated statistical rules as determining factor in assessors' judgement and actions
- **Mechanical Prediction –** application of computer algorithms together with statistical rules and probabilities to generate findings and recommendations

Levels of Interpretation

Level I – there is a minimal amount of any sort of interpretation
- minimal concern with intervening processes
- data are primarily treated in a sampling or correlate way
- no concern with underlying constructs
- found in large-scale selection testing
- for psychometric approaches

Level II
a. *Descriptive Generalizations*
b. *Hypothetical Construct:* the assumption of an inner state which goes logically beyond the description of visible behavior

Level III – the effort to develop a coherent and inclusive theory of the individual life or a "working image" of the patient
- the clinician attempts full scale exploration of the individual's personality, psychosocial situation, and developmental history

- **Extra-Test Behavior –** observations made by an examiner regarding what the examinee does and how the examinee reacts during the course of testing that are indirectly related to the test's specific content but of possible significance to interpretation

Parties in Psychological Assessment

1. Test Author/Developer – creates the tests or other methods of assessment

2. Test Publishers – they publish, market, sell, and control the distribution of tests

3. Test Reviewers – prepare evaluative critiques based on the technical and practical aspects of the tests

4. Test Users – uses the test of assessment

5. Test Takers – those who take the tests

6. Test Sponsors – institutions or government who contract test developers for various testing services

7. Society

- **Test Battery –** selection of tests and assessment procedures typically composed of tests designed to measure different variables but having a common objective

Assumptions about Psychological Testing and Assessment

Assumption 1: Psychological Traits and States Exist

- **Trait –** any distinguishable, relatively enduring way in which one individual varies from another

- Permit people predict the present from the past
- Characteristic patterns of thinking, feeling, and behaving that generalize across similar situations, differ systematically between individuals, and remain rather stable across time
- *Psychological Trait* – **i**ntelligence, specific intellectual abilities, cognitive style, adjustment, interests, attitudes, sexual orientation and preferences, psychopathology, etc.

- **States –** distinguish one person from another but are relatively less enduring

- Characteristic pattern of thinking, feeling, and behaving in a concrete situation at a specific moment in time
- Identify those behaviors that can be controlled by manipulating the situation

- Psychological Traits exists as construct

- *Construct:* an informed, scientific concept developed or constructed to explain a behavior, inferred from overt behavior
- *Overt Behavior:* an observable action or the product of an observable action

- Trait is not expected to be manifested in behavior 100% of the time
- Whether a trait manifests itself in observable behavior, and to what degree it manifests, is presumed to depend not only on the strength of the trait in the individual but also on the nature of the action (situation-dependent)
- Context within which behavior occurs also plays a role in helping us select appropriate trait terms for observed behaviors
- Definition of trait and state also refer to a way in which one individual varies from another
- Assessors may make comparisons among people who, because of their membership in some group or for any number of other reasons, are decidedly not average

Source: Cohen & Swerdlik (2018), Kaplan & Saccuzzo (2018), Groth & Wright (2016), Psych Pearls

Assumption 2: Psychological Traits and States can be Quantified and Measured

- Once the trait, state or other construct has been defined to be measured, a test developer consider the types of item content that would provide insight to it, to gauge the strength of that trait
- Measuring traits and states means of a test entails developing not only appropriate tests items but also appropriate ways to score the test and interpret the results
- **Cumulative Scoring** – assumption that the more the testtaker responds in a particular direction keyed by the test manual as correct or consistent with a particular trait, the higher that testtaker is presumed to be on the targeted ability or trait

Assumption 3: Test-Rlated Behavior Predicts Non-Test-Related Behavior

- The tasks in some tests mimics the actual behaviors that the test user is attempting to understand
- Such tests only yield a sample of the behavior that can be expected to be emitted under nontest conditions

Assumption 4: Test and Other Measurement Techniques have strengths and weaknesses

- Competent test users understand and appreciate the limitations of the test they use as well as how those limitations might be compensated for by data from other sources

Assumption 5: Various Sources of Error are part of the Assessment Process

- **Error** – refers to something that is more than expected; it is component of the measurement process
  - Refers to a long-standing assumption that factors other than what a test attempts to measure will influence performance on the test
  - **Error Variance** – the component of a test score attributable to sources other than the trait or ability measured
- Potential Sources of error variance:
  1. Assessors
  2. Measuring Instruments
  3. Random errors such as luck
- **Classical Test Theory** – each testtaker has true score on a test that would be obtained but for the action of measurement error

Assumption 6: Testing and Assessment can be conducted in a Fair and Unbiased Manner

- Despite best efforts of many professionals, fairness-related questions and problems do occasionally rise

In al questions about tests with regards to fairness, it is important to keep in mind that tests are tools ⊢they can be used properly or improperly

Assumption 7: Testing and Assessment Benefit Society

- Considering the many critical decisions that are based on testing and assessment procedures, we can readily appreciate the need for tests

Reliability

- **Reliability** – dependability or consistency of the instrument or scores obtained by the same person when re-examined with the same test on different occasions, or with different sets of equivalent items
  - Test may be reliable in one context, but unreliable in another
  - Estimate the range of possible random fluctuations that can be expected in an individual's score
  - Free from errors
  - More number of items = higher reliability
  - Minimizing error
  - Using only representative sample to obtain an observed score
  - True score cannot be found
  - *Reliability Coefficient:* index of reliability, a proportion that indicates the ratio between the true score variance on a test and the total variance
- **Classical Test Theory (True Score Theory)** – score on an ability tests is presumed to reflect not only the testtaker's true score on the ability being measured but also the error
  - *Error:* refers to the component of the observed test score that does not have to do with the testtaker's ability
  - Errors of measurement are random

$$X = T + E$$

X = observed score
T = true score
E = Error

Source: Cohen & Swerdlik (2018), Kaplan & Saccuzzo (2018), Groth & Wright (2016), Psych Pearls

- When you average all the observed scores obtained over a period of time, then the result would be closest to the true score
- *The greater number of items, the higher the reliability*
- Factors the contribute to consistency: *stable attributes*
- Factors that contribute to inconsistency: *characteristics* of the individual, test, or situation, which have nothing to do with the attribute being measured, but still affect the scores

- Goals of Reliability:
  - ✓ Estimate errors
  - ✓ Devise techniques to improve testing and reduce errors
- **Variance –** useful in describing sources of test score variability
  - *True Variance:* variance from true differences
  - *Error Variance:* variance from irrelevant random sources

Measurement Error – all of the factors associated with the process of measuring some variable, other than the variable being measured
- difference between the observed score and the true score
- *Positive:* can increase one's score
- *Negative:* decrease one's score
- Sources of Error Variance:

a. Item Sampling/Content Sampling: refer to variation among items within a test as well as to variation among items between tests
- The extent to which testtaker's score is affected by the content sampled on a test and by the way the content is sampled is a source of error variance

b. Test Administration: testtaker's motivation or attention, environment, etc.

c. Test Scoring and Interpretation: may employ objective-type items amenable to computer scoring of well-documented reliability

Random Error – source of error in measuring a targeted variable caused by unpredictable fluctuations and inconsistencies of other variables in measurement process (e.g., noise, temperature, weather)

Systematic Error – source of error in a measuring a variable that is typically constant or proportionate to what is presumed to be the true values of the variable being measured
- has consistent effect on the true score
- SD does not change, the mean does

- Reliability refers to the proportion of total variance attributed to true variance
- The greater the proportion of the total variance attributed to true variance, the more reliable the test
- Error variance may increase or decrease a test score by varying amounts, consistency of test score, and thus, the reliability can be affected

True Score Formula:

$$\text{True score} = R_{xx}(x - \bar{x}) + \bar{x}$$

Wherein, R_{xx} = correlation coefficient
x = obtained score
$\bar{x}$ = mean

Test-Retest Reliability

Error: *Time Sampling*
- *time sampling reliability*
- an estimate of reliability obtained by correlating pairs of scores from the same people on two different administrations of the test
- appropriate when evaluating the reliability of a test that purports to measure an *enduring* and *stable attribute* such as personality trait
- established by comparing the scores obtained from two successive measurements of the same individuals and calculating a correlated between the two set of scores
- the longer the time passes, the greater likelihood that the reliability coefficient would be insignificant
- *Carryover Effects:* happened when the test-retest interval is short, wherein the second test is influenced by the first test because they remember or practiced the previous test = inflated correlation/overestimation of reliability
- *Practice Effect:* scores on the second session are higher due to their experience of the first session of testing
- *Test Sophistication:* items are remembered by the test takers especially the difficult ones/items that we got highlight confused
- *Test Wiseness:* might inflate the abilities of test takers
- test-retest with longer interval might be affected of other extreme factors, thus, resulting to low correlation
- lower correlation = poor reliability

Source: Cohen & Swerdlik (2018), Kaplan & Saccuzzo (2018), Groth & Wright (2016), Psych Pearls

- *Mortality:* problems in absences in second session (just remove the first tests of the absents)
- intelligence, motor test, traits, reaction time (stable traits)
- *Coefficient of Stability*
- statistical tool: *Pearson R, Spearman Rho*

Parallel Forms/Alternate Forms Reliability

Error: *Item Sampling* (Immediate), *Item Sampling changes over time* (delayed)
- established when at least two different versions of the test yield almost the same scores
- has the most universal applicability
- *Parallel Forms:* each form of the test, the means, and the error variances, are EQUAL; same items, different positionings/numberings
- true score must be the same for two test
- *Alternate Forms:* simply different version of a test that has been constructed so as to be parallel
- test should contain the same number of items and the items should be expressed in the same form and should cover the same type of content; range and difficulty must also be equal
- if there is a test leakage, use the form that is not mostly administered
- *Counterbalancing:* technique to avoid carryover effects for parallel forms, by using different sequence for groups (e.g., G1 – listen to song before counseling, G2 – counseling first, before listening to the song)
- can be administered on the same day or different time
- most rigorous and burdensome, since test developers create two forms of the test
- main problem: difference between the two test
- test scores may be affected by motivation, fatigue, or intervening events
- means and the variances of the observed scores must be equal for two forms
- Statistical Tool: *Pearson R* or *Spearman Rho*

Internal Consistency (Inter-Item Reliability)

Error: *Item Sampling Homogeneity*
- used when tests are administered once
- consistency among items within the test
- measures the internal consistency of the test which is the degree to which each item measures the same construct
- measurement for *unstable* traits
- if all items measure the same construct, then it has a good internal consistency
- useful in assessing Homogeneity
- *Homogeneity:* if a test contains items that measure a single trait (unifactorial)
- *Heterogeneity:* degree to which a test measures different factors (more than one factor/trait)
- more homogenous = higher inter-item consistency
- *KR-20:* used for inter-item consistency of dichotomous items (intelligence tests, personality tests with yes or no options, multiple choice), unequal variances, dichotomous scored
- *KR-21:* if all the items have the same degree of difficulty (speed tests), equal variances, dichotomous scored
- *Cronbach's Coefficient Alpha:* used when two halves of the test have unequal variances and on tests containing non-dichotomous items, unequal variances
- *Average Proportional Distance:* measure used to evaluate internal consistencies of a test that focuses on the degree of differences that exists between item scores

Split-Half Reliability

Error: *Item sample: Nature of Split*
- *Split Half Reliability:* obtained by correlating two pairs of scores obtained from equivalent halves of a single test administered ONCE
- useful when it is impractical or undesirable to assess reliability with two tests or to administer a test twice
- cannot just divide the items in the middle because it might spuriously raise or lower the reliability coefficient, so just randomly assign items or assign odd-numbered items to one half and even-numbered items to the other half
- *Spearman-Brown Formula:* allows a test developer of user to estimate internal consistency reliability from a correlation of two halves of a test, if each half had been the length of the whole test and have the equal variances
- *Spearman-Brown Prophecy Formula:* estimates how many more items are needed in order to achieve the target reliability
- multiply the estimate to the original number of items
- *Rulon's Formula:* counterpart of spearman-brown formula, which is the ratio of the variance of difference between the odd and even splits and the variance of the total, combined odd-even, score
- if the reliability of the original test is relatively low, then developer could create new items, clarify test instructions, or simplifying the scoring rules
- equal variances, dichotomous scored
- Statistical Tool: *Pearson R* or *Spearman Rho*

Inter-Scorer Reliability

Source: Cohen & Swerdlik (2018), Kaplan & Saccuzzo (2018), Groth & Wright (2016), Psych Pearls

Error: *Scorer Differences*
- the degree of agreement or consistency between two or more scorers with regard to a particular measure'
- evaluated by calculating the percentage of times that two individuals assign the same scores to the performance of the examinees
- a variation is to have two different examiners test the same client using the same test and then to determine how close their scores or ratings of the person are
- used for coding nonbehavioral behavior
- observer differences
- *Fleiss Kappa:* determine the level between TWO or MORE raters when the method of assessment is measured on CATEGORICAL SCALE
- *Cohen's Kappa:* two raters only
- *Krippendorff's Alpha:* two or more rater, based on observed disagreement corrected for disagreement expected by chance

- Tests designed to measure one factor (Homogenous) are expected to have high degree of internal consistency and vice versa
- **Dynamic** – trait, state, or ability presumed to be ever-changing as a function of situational and cognitive experience
- **Static** – barely changing or relatively unchanging
- **Restriction of range** or **Restriction of variance** – if the variance of either variable in a correlational analysis is restricted by the sampling procedure used, then the resulting correlation coefficient tends to be lower
- **Power Tests** – when time limit is long enough to allow test takers to attempt all times
- **Speed Tests** – generally contains items of uniform level of difficulty with time limit
  - Reliability should be based on performance from two independent testing periods using test-retest and alternate-forms or split-half-reliability
- **Criterion-Referenced Tests** – designed to provide an indication of where a testtaker stands with respect to some variable or criterion
  - As individual differences decrease, a traditional measure of reliability would also decrease, regardless of the stability of individual performance
- **Classical Test Theory** – everyone has a "true score" on test
  - *True Score:* genuinely reflects an individual's ability level as measured by a particular test
  - Random Error
- **Domain Sampling Theory** – estimate the extent to which specific sources of variation under defined conditions are contributing to the test scores
  - Considers problem created by using a limited number of items to represent a larger and more complicated construct
  - Test reliability is conceived of as an objective measure of how precisely the test score assesses the domain from which the test draws a sample
  - *Generalizability Theory:* based on the idea that a person's test scores vary from testing to testing because of the variables in the testing situations
  - *Universe:* test situation
  - *Facets:* number of items in the test, amount of review, and the purpose of test administration
  - According to Generalizability Theory, given the exact same conditions of all the facets in the universe, the exact same test score should be obtained (Universe score)
  - *Decision Study:* developers examine the usefulness of test scores in helping the test user make decisions
  - Systematic Error
- **Item Response Theory** – the probability that a person with X ability will be able to perform at a level of Y in a test
  - Focus: item difficulty
  - **Latent-Trait Theory**
  - a system of assumption about measurement and the extent to which item measures the trait
  - The computer is used to focus on the range of item difficulty that helps assess an individual's ability level
  - If you got several easy items correct, the computer will them move to more difficult items
  - *Difficulty:* attribute of not being easily accomplished, solved, or comprehended
  - *Discrimination:* degree to which an item differentiates among people with higher or lower levels of the trait, ability etc.
  - *Dichotomous:* can be answered with only one of two alternative responses
  - *Polytomous:* 3 or more alternative responses
- **Standard Error of Measurement** – provide a measure of the precision of an observed test score
  - Standard deviation of errors as the basic measure of error
  - Index of the amount of inconsistent or the amount of the *expected* error in an individual's score
  - Allows to quantify the extent to which a test provide accurate scores

Source: Cohen & Swerdlik (2018), Kaplan & Saccuzzo (2018), Groth & Wright (2016), Psych Pearls

- Provides an estimate of the amount of error inherent in an observed score or measurement
- Higher reliability, lower SEM
- Used to estimate or infer the extent to which an observed score deviates from a true score
- **Standard Error of a Score**
- *Confidence Interval:* a range or band of test scores that is likely to contain true scores

- **Standard Error of the Difference** – can aid a test user in determining how large a difference should be before it is considered statistically significant
- **Standard Error of Estimate** – refers to the standard error of the difference between the predicted and observed values
- **Confidence Interval** – a range of and of test score that is likely to contain true score
  - Tells us the relative ability of the true score within the specified range and confidence level
  - The larger the range, the higher the confidence
- If the reliability is low, you can increase the number of items or use factor analysis and item analysis to increase internal consistency
- **Reliability Estimates** – nature of the test will often determine the reliability metric
  a) Homogenous (unifactor) or heterogeneous (multifactor)
  b) Dynamic (unstable) or static (stable)
  c) Range of scores is restricted or not
  d) Speed Test or Power Test
  e) Criterion or non-Criterion
- **Test Sensitivity** – detects true positive
- **Test Specificity** – detects true negative
- **Base Rate** – proportion of the population that actually possess the characteristic of interest
- **Selection ratio** – no. of hired candidates compared to the no. of applicants

$$Selection\ ratio = \frac{Number\ of\ hired\ candidates}{Total\ number\ of\ candidates}$$

- **Four Possible Hit and Miss Outcomes**
  1. **True Positives (Sensitivity)** – predict success that does occur
  2. **True Negatives (Specificity)** – predict failure that does occur
  3. **False Positive (Type 1)** – success does not occur
  4. **False Negative (Type 2)** – predicted failure but succeed

Validity

- **Validity** – a judgment or estimate of how well a test measures what it supposed to measure
  - Evidence about the appropriateness of inferences drawn from test scores
  - Degree to which the measurement procedure measures the variables to measure
  - **Inferences** – logical result or deduction
  - May diminish as the culture or times change
  - ✓ Predicts future performance
  - ✓ Measures appropriate domain
  - ✓ Measures appropriate characteristics
- **Validation** – the process of gathering and evaluating evidence about validity
- **Validation Studies** – yield insights regarding a particular population of testtakers as compared to the norming sample described in a test manual
- **Internal Validity** – degree of control among variables in the study (increased through random assignment)
- **External Validity** – generalizability of the research results (increased through random selection)
- **Conceptual Validity** – focuses on individual with their unique histories and behaviors
  - Means of evaluating and integrating test data so that the clinician's conclusions make accurate statements about the examinee
- **Face Validity** – a test appears to measure to the person being tested than to what the test actually measures

Content Validity
- describes a judgement of how adequately a test samples behavior representative of the universe of behavior that the test was designed to sample - representativeness and relevance of the assessment instrument to the construct being measured - when the proportion of the material covered by the test approximates the proportion of material covered in the course

Source: Cohen & Swerdlik (2018), Kaplan & Saccuzzo (2018), Groth & Wright (2016), Psych Pearls

- *Test Blueprint:* a plan regarding the types of information to be covered by the items, the no. of items tapping each area of coverage, the organization of the items, and so forth
- more logical than statistical
- concerned with the extent to which the test is representative of defined body of content consisting the topics and processes
- panel of experts can review the test items and rate them in terms of how closely they match the objective or domain specification
- examine if items are essential, useful and necessary
- *Construct underrepresentation:* failure to capture important components of a construct
- *Construct-irrelevant variance:* happens when scores are influenced by factors irrelevant to the construct
- *Lawshe:* developed the formula of Content Validity Ratio

$$CVR = \frac{N_e - (\frac{N}{2})}{N/2}$$

- If the CVI is low, it is recommended to remove or modify the items that have low CVR values to improve the overall content validity of the test
- *Zero CVR:* exactly half of the experts rate the item as essential

Criterion Validity

- more statistical than logical
- a judgement of how adequately a test score can be used to infer an individual's most probable standing on some measure of interest – the measure of interest being criterion
- *Criterion:* standard on which a judgement or decision may be made
- Characteristics: relevant, valid, uncontaminated
- *Criterion Contamination:* occurs when the criterion measure includes aspects of performance that are not part of the job or when the measure is affected by "construct-irrelevant" (Messick, 1989) factors that are not part of the criterion construct

1. *Concurrent Validity:* If the test scores obtained at about the same time as the criterion measures are obtained; economically efficient
2. *Predictive Validity:* measures of the relationship between test scores and a criterion measure obtained at a future time

- *Incremental Validity:* the degree to which an additional predictor explains something about the criterion measure that is not explained by predictors already in use; used to improve the domain

Construct Validity (Umbrella Validity)

- covers all types of validity
- logical and statistical
- judgement about the appropriateness of inferences drawn from test scores regarding individual standing on variable called construct
- *Construct:* an informed, scientific idea developed or hypothesized to describe or explain behavior; unobservable, presupposed traits that may invoke to describe test behavior or criterion performance
- One way a test developer can improve the homogeneity of a test containing dichotomous items is by eliminating items that do not show significant correlation coefficients with total test scores
- If it is an academic test and high scorers on the entire test for some reason tended to get that particular item wrong while low scorers got it right, then the item is obviously not a good one
- Some constructs lend themselves more readily than others to predictions of change over time
- *Method of Contrasted Groups:* demonstrate that scores on the test vary in a predictable way as a function of membership in a group
- If a test is a valid measure of a particular construct, then the scores from the group of people who does not have that construct would have different test scores than those who really possesses that construct
- *Convergent Evidence:* if scores on the test undergoing construct validation tend to highly correlated with another established, validated test that measures the same construct
- *Discriminant Evidence:* a validity coefficient showing little relationship between test scores and/or other variables with which scores on the test being construct-validated should not be correlated
- test is homogenous
- test score increases or decreases as a function of age, passage of time, or experimental manipulation
- pretest-posttest differences
- scores differ from groups
- scores correlated with scores on other test in accordance to what is predicted
- *Sensitivity:* percentage of true positives
- *Specificity:* percentage of true negatives

- **Multitrait-multimethod Matrix** – useful for examining both convergent and discriminant validity evidence
  - *Multitrait:* two or more traits
  - *Multimethod:* two or more methods

Source: Cohen & Swerdlik (2018), Kaplan & Saccuzzo (2018), Groth & Wright (2016), Psych Pearls

- The matrix or table that results from correlating variables within and between methods

- **Factor Analysis** – designed to identify factors or specific variables that are typically attributes, characteristics, or dimensions on which people may differ
  - Developed by Charles Spearman
  - Employed as data reduction method
  - Used to study the interrelationships among set of variables
  - Identify the factor or factors in common between test scores on subscales within a particular test
  - *Explanatory FA:* estimating or extracting factors; deciding how many factors must be retained
  - *Confirmatory FA:* researchers test the degree to which a hypothetical model fits the actual data
  - *Factor Loading:* conveys info about the extent to which the factor determines the test score or scores
  - can be used to obtain both convergent and discriminant validity
- **Cross-Validation** – revalidation of the test to a criterion based on another group different from the original group form which the test was validated
  - *Validity Shrinkage:* decrease in validity after cross-validation
  - *Co-Validation:* validation of more than one test from the same group
  - *Co-Norming:* norming more than one test from the same group
- **Bias** – factor inherent in a test that systematically prevents accurate, impartial measurement
  - Prejudice, preferential treatment
  - Prevention during test dev through a procedure called **Estimated True Score Transformation**
- **Rating** – numerical or verbal judgement that places a person or an attribute along a continuum identified by a scale of numerical or word descriptors known as **Rating Scale**
  - *Rating Error:* intentional or unintentional misuse of the scale
  - *Leniency Error:* rater is lenient in scoring (Generosity Error)
  - *Severity Error:* rater is strict in scoring
  - *Central Tendency Error:* rater's rating would tend to cluster in the middle of the rating scale
  - One way to overcome rating errors is to use rankings
  - *Halo Effect:* tendency to give high score due to failure to discriminate among conceptually distinct and potentially independent aspects of a ratee's behavior
- **Fairness** – the extent to which a test is used in an impartial, just, and equitable way
- Attempting to define the validity of the test will be futile if the test is NOT reliable

Utility

- **Utility** – usefulness or practical value of testing to improve efficiency
- Can tell us something about the practical value of the information derived from scores on the test
- Helps us make better decisions
- Higher criterion-related validity = higher utility
- One of the most basic elements in utility analysis is financial cost of the selection device
- **Cost** – disadvantages, losses, or expenses both economic and noneconomic terms
- **Benefit** – profits, gains or advantages
- The cost of test administration can be well worth it if the results is certain noneconomic benefits
- **Utility Analysis** – family of techniques that entail a cost-benefit analysis designed to yield information relevant to a decision about the usefulness and/or practical value of a tool of assessment
- **Expectancy table** – provide an indication that a testtaker will score within some interval of scores on a criterion measure – passing, acceptable, failing
- Might indicate future behaviors, then if successful, the test is working as it should
- **Taylor-Russel Tables** – provide an estimate of the extent to which inclusion of a particular test in the selection system will improve selection
- **Selection Ratio** – numerical value that reflects the relationship between the number of people to be hired and the number of people available to be hired

$$Selection\ ratio = \frac{Number\ of\ hired\ candidates}{Total\ number\ of\ candidates}$$

- **Base Rate** – percentage of people hired under the existing system for a particular position
- One limitation of Taylor-Russel Tables is that the relationship between the predictor (test) and criterion must be linear
- **Naylor-Shine Tables** – entails obtaining the difference between the means of the selected and unselected groups to derive an index of what the test is adding to already established procedures
- **Brogden-Cronbach-Gleser Formula** – used to calculate the dollar amount of a utility gain resulting from the use of a particular selection instrument

Source: Cohen & Swerdlik (2018), Kaplan & Saccuzzo (2018), Groth & Wright (2016), Psych Pearls

- **Utility Gain** – estimate of the benefit of using a particular test
- **Productivity Gains** – an estimated increase in work output
- High performing applicants may have been offered in other companies as well
- The more complex the job, the more people differ on how well or poorly they do that job
- **Cut Score** – reference point derived as a result of a judgement and used to divide a set of data into two or more classifications

Relative Cut Score – reference point based on norm-related considerations (norm-referenced), not fixed per se
Fixed Cut Scores – set with reference to a judgement concerning minimum level of proficiency required; e.g., Board Exams
Multiple Cut Scores – refers to the use of two or more cut scores with reference to one predictor for the purpose of categorization
Multiple Hurdle – multi-stage selection process, a cut score is in place for each predictor
Compensatory Model of Selection – assumption that high scores on one attribute can compensate for lower scores

- **Angoff Method** – setting fixed cut scores
  - low interrater reliability
  - requires expert judges to discuss the issues involved in determining a pass mark and to evaluate the examination by using a well-defined and rational procedure
- **Known Groups Method** – collection of data on the predictor of interest from group known to possess and not possess a trait of interest
  - The determination of where to set cutoff score is inherently affected by the composition of contrasting groups
- **IRT-Based Methods** – cut scores are typically set based on testtaker's performance across all the items on the test
  - *Item-Mapping Method*: arrangement of items in histogram, with each column containing items with deemed to be equivalent value
  - *Bookmark Method*: expert places "bookmark" between the two pages that are deemed to separate testtakers who have acquired the minimal knowledge, skills, and/or abilities from those who are not
- **Method of Predictive Yield** – took into account the number of positions to be filled, projections regarding the likelihood of offer acceptance, and the distribution of applicant scores
- **Discriminant Analysis** – shed light on the relationship between identified variables and two naturally occurring groups
  - used by the researcher to analyze the research data when the criterion or the dependent variable is categorical and the predictor or the independent variable is interval in nature
  - enables the researcher to examine whether significant differences exist among the groups, in terms of the predictor variables
  - identify two groups of people who represent two distinct categories of some trait

Reason for accepting or rejecting instruments and tools based on Psychometric Properties

Reliability

Reliability coefficient value	Interpretation
.90 and up	excellent
.80 - .89	good
.70 - .79	adequate
below .70	may have limited applicability

- Basic Research = 0.70 to 0.90
- Clinical Setting = 0.90 to 0.95

Validity

Validity coefficient value	Interpretation
above .35	very beneficial
.21 - .35	likely to be useful
.11 - .20	depends on circumstances
below .11	unlikely to be useful

Source: Cohen & Swerdlik (2018), Kaplan & Saccuzzo (2018), Groth & Wright (2016), Psych Pearls

Item Difficulty

Item Difficulty Range	Level of Difficulty
0.0-0.19	Very difficult
0.20-0.39	Difficult
0.40-0.60	Average/moderately difficult
0.61-0.79	Easy
0.80-1.0	Very easy

Item Discrimination

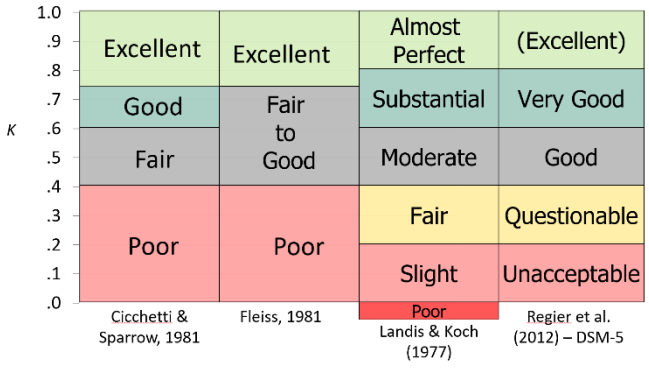
Range	Verbal Description
.40 & above	Very Good Item
.30 - .39	Good Item
.20 - .29	Fair Item
.09 - .19	Poor Item

- The optimal boundary lines for the "upper" and "lower" areas of distribution of scores will demarcate the upper and lower 27% of distribution of scores
  - 27% if normal
  - 33% if platykurtic
- ***0.27/n,*** wherein n = no. of students

Inter-Item Reliability Index

Cronbach's alpha	Internal consistency
α ≥ 0.9	Excellent
0.9 > α ≥ 0.8	Good
0.8 > α ≥ 0.7	Acceptable
0.7 > α ≥ 0.6	Questionable
0.6 > α ≥ 0.5	Poor
0.5 > α	Unacceptable

Interrater Reliability Coefficient



- 0 means 0% of the variance in the scores assigned by the scorers was attributed to true differences and 100% to error

P-Value

- P-Value ≤ ∞, reject null hypothesis
- P-Value ≥ ∞, accept null hypothesis

p-value	Evidence against H_0
$p > 0.10$	Weak or no evidence
$0.05 < p \leq 0.10$	Moderate evidence
$0.01 < p \leq 0.05$	Strong evidence
$p \leq 0.01$	Very strong evidence

Research Methods and Statistics (20)

Statistics Applied in Research Studies on tests and Tests Development

Measures of Central Tendency - statistics that indicates the *average or midmost* score between the extreme scores in a distribution
- Goal: Identify the most typical or representative of entire group
- *Measures of Central Location*

Mean	- the average of all the raw scores - Equal to the sum of the observations divided by the number of observations - *Interval* and *ratio data* (when normal distribution) - Point of least squares - Balance point for the distribution - susceptible to outliers
Median	– the middle score of the distribution - Ordinal, Interval, Ratio - for extreme scores, use median - Identical for sample and population - Also used when there has an unknown or undetermined score - Used in "open-ended" categories (e.g., 5 or more, more than 8, at least 10)

	- For *ordinal* data - if the distribution is skewed for ratio/interval data, use median
Mode	- most frequently occurring score in the distribution - *Bimodal Distribution:* if there are two scores that occur with highest frequency - Not commonly used - Useful in analyses of qualitative or verbal nature - For *nominal* scales, discrete variables - Value of the mode gives an indication of the shape of the distribution as well as a measure of central tendency
Measures of Spread or Variability – statistics that describe the amount of variation in a distribution - gives idea of how well the measure of central tendency represent the data - large spread of values means large differences between individual scores	
Range	- equal to the difference between highest and the lowest score - Provides a quick but gross description of the spread of scores - When its value is based on extreme scores of the distribution, the resulting description of variation may be understated or overstated
Interquartile Range	- difference between Q1 and Q2
Semi-Quartile Range	- interquartile range divided by 2
Standard Deviation	- approximation of the average deviation around the mean - gives detail of how much above or below a score to the mean
	- equal to the square root of the average squared deviations about the mean - Equal to the square root of the variance - Distance from the mean
Variance	- equal to the arithmetic mean of the squares of the differences between the scores in a distribution and their mean - average squared deviation around the mean
Measures of Location	
Percentile or Percentile Rank	- not linearly transformable, converged at the middle and the outer ends show large interval - expressed in terms of the percentage of persons in the standardization sample who fall below a given score - indicates the individual's relative position in the standardization sample - essential in creating normalized standardized scores
Quartile	- dividing points between the four quarters in the distribution - Specific point - *Quarter:* refers to an interval
Decile/STEN	- divide into 10 equal parts - a measure of the asymmetry of the probability distribution of a real-valued random about its mean
Correlation	
Pearson R	- interval/ratio + interval/ratio

Source: Cohen & Swerdlik (2018), Kaplan & Saccuzzo (2018), Groth & Wright (2016), Psych Pearls

Spearman Rho	- ordinal + ordinal
Biserial	- artificial Dichotomous + interval/ratio
Point Biserial	- true dichotomous + interval/ratio
Phi Coefficient	- nominal (true dic) + nominal (true/artificial dic.)
Tetrachoric	- Art. Dichotomous + Art. Dichotomous
Kendall's	- 3 or more ordinal/rank
Rank Biserial	- nominal + ordinal
Differences	
T-test Independent (Unpaired T-test)	- two separate groups, random assignment - e.g., blood pressure of male and female grad students
T-Test Dependent (Paired T-test)	- one group, two scores - e.g., blood pressure before and after the lecture of Grad students
One-Way ANOVA	- 3 or more groups, tested once - e.g., people in different socio-economic status and the differences of their salaries
One-Way Repeated Measures	- 1 group, measured at least 3 times - e.g., measuring the focus level of board reviewers during morning, afternoon, and night sessions of review
Two-Way ANOVA	- 3 or more groups, tested for 2 variables - e.g., people in different socio-economic status and the differences of their salaries and their eating habits
ANCOVA	- used when you need to control for an additional variable which may be influencing the relationship between your independent and dependent variable
ANOVA Mixed Design	- 2 or more groups, measured more than 3 times - e.g., Young Adults, Middle Adults, and Old Adults' blood pressure is measured during breakfast, lunch, and dinner
MANOVA	- used to test the differences between the means of multiple dependent variables across two or more independent groups
Non-Parametric Tests	
Mann Whitney U Test	- t-test independent
Wilcoxon Signed Rank Test	- t-test dependent
Kruskal-Wallis H Test	- one-way/two-way ANOVA
Friedman Test	- ANOVA repeated measures
Lambda	- for 2 groups of nominal data
Chi-Square	
Goodness of Fit	- used to measure differences and involves nominal data and only one variable with 2 or more categories
Test of Independence	- used to measure correlation and involves nominal data and two variables with two or more categories
Regression – used when one wants to provide framework of prediction on the basis of one factor in order to predict the probable value of another factor	
Linear Regression of Y on X	- $Y = a + bX$ - Used to predict the unknown value of variable Y when value of variable X is known
Linear Regression of X on Y	- $X = c + dY$ - Used to predict the unknown value of variable X using the known variable Y

Source: Cohen & Swerdlik (2018), Kaplan & Saccuzzo (2018), Groth & Wright (2016), Psych Pearls

Test for Normality – data are drawn from a population that has normal distribution

Kolmogorov-Smirnov	More than 50 sample size
Shapiro-Wilk	Less than 50 sample size

	No. Times DV is measured	Number of Groups	scale of measurement of DV
1. t-test independent means	1	2	Interval/Ratio
2. t-test dependent means	2	1	Interval/Ratio
3. ANOVA 1 way	1	>2	Interval/Ratio
4. ANOVA Repeated Measures	>2	1	Interval/Ratio
5. Mann Whitney U Test	1	2	Ordinal
6. Wilcoxon Signed Rank Test	2	1	Ordinal
7. Kruskal Wallis H Test	1	>2	Ordinal
8. Friedman Test	>2	1	Ordinal

- **True Dichotomy** – dichotomy in which there are only fixed possible categories
- **Artificial Dichotomy** - dichotomy in which there are other possibilities in a certain category
- **Levene's Test** – used to test if k samples have equal variance
  - P-Value > 0.05, variance not significantly different from each other (therefore, homogenous)
  - P-Value < 0.05, there is significant difference between variance
  - Less sensitive to departures from normality

H_0: Groups have equal variances

H_1: Groups have different variances

- **Bartlett's Test** – also used to test if k samples have equal variance but much sensitive than Levene Test

Methods and Statistics used in Research Studies and Test Construction

Test Development

- **Test Development** – an umbrella term for all that goes into the process of creating a test

I. Test Conceptualization – brainstorming of ideas about what kind of test a developer wants to publish
- stage wherein the ff. is determined: construct, goal, user, taker, administration, format, response, benefits, costs, interpretation
- determines whether the test would be norm-referenced or criterion-referenced
- Pilot Work/Pilot Study/Pilot Research – preliminary research surrounding the creation of a prototype of the test
- Attempts to determine how best to measure a targeted construct
- Entail lit reviews and experimentation, creation, revision, and deletion of preliminary items

II. Test Construction – stage in the process that entails writing test items, revisions, formatting, setting scoring rules
- it is not good to create an item that contains numerous ideas
- *Item Pool:* reservoir or well from which the items will or will not be drawn for the final version of the test
- *Item Banks:* relatively large and easily accessible collection of test questions
- *Computerized Adaptive Testing:* refers to an interactive, computer administered test-taking process wherein items presented to the testtaker are based in part on the testtaker's performance on previous items
- The test administered may be different for each testtaker, depending on the test performance on the items presented
- Reduces floor and ceiling effects
- *Floor Effects:* occurs when there is some lower limit on a survey or questionnaire and a large percentage of respondents score near this lower limit (testtakers have low scores)
- *Ceiling Effects:* occurs when there is some upper limit on a survey or questionnaire and a large percentage of respondents score near this upper limit (testtakers have high scores)
- *Item Branching:* ability of the computer to tailor the content and order of presentation of items on the basis of responses to previous items
- *Item Format:* form, plan, structure, arrangement, and layout of individual test items
- *Dichotomous Format:* offers two alternatives for each item
- *Polychotomous Format:* each item has more than two alternatives
- *Category Format:* a format where respondents are asked to rate a construct

1. **Checklist** – subject receives a longlist of adjectives and indicates whether each one if characteristic of himself or herself
2. **Guttman Scale** – items are arranged from weaker to stronger expressions of attitude, belief, or feelings
- *Selected-Response Format:* require testtakers to select response from a set of alternative responses

1. **Multiple Choice** - Has three elements: *stem* (question), a correct option, and several incorrect alternatives (*distractors or foils*), Should've one correct answer, has grammatically parallel alternatives, similar length, alternatives that fit grammatically with

the stem, avoid ridiculous distractors, not excessively long, "all of the above", "none of the above" (25%)
- *Effective Distractors:* a distractor that was chosen equally by both high and low performing groups that enhances the consistency of test results
- Good distractors has been chosen frequently by low scorers
- *Ineffective Distractors:* may hurt the reliability of the test because they are time consuming to read and can limit the no. of good items
- *Cute Distractors:* less likely to be chosen, may affect the reliability of the test bec the testtakers may guess from the remaining options

2. **Matching Item -** Test taker is presented with two columns: Premises and Responses
3. **Binary Choice -** Usually takes the form of a sentence that requires the testtaker to indicate whether the statement is or is not a fact (50%)
- *Constructed-Response Format:* requires testtakers to supply or to create the correct answer, not merely selecting it

1. **Completion Item -** Requires the examinee to provide a word or phrase that completes a sentence
2. **Short-Answer -** Should be written clearly enough that the testtaker can respond succinctly, with short answer
3. **Essay** – allows creative integration and expression of the material
- *Scaling:* process of setting rules for assigning numbers in measurement

Primary Scales of Measurement

1. **Nominal** - involve classification or categorization based on one or more distinguishing characteristics
- Label and categorize observations but do not make any quantitative distinctions between observations
- mode

2**. Ordinal -** rank ordering on some characteristics is also permissible
- median

3. **Ratio -** contains equal intervals, has no absolute zero point (even negative values have interpretation to it)
- Zero value does not mean it represents none

4. **Interval - -** has true zero point (if the score is zero, it means none/null)
- Easiest to manipulate

Comparative Scales of Measurement

1. **Paired Comparison -** produces ordinal data by presenting with pairs of two stimuli which they are asked to compare
- respondent is presented with two objects at a time and asked to select one object according to some criterion

2. **Rank Order** – respondents are presented with several items simultaneously and asked to rank them in order or priority
3. **Constant Sum** – respondents are asked to allocate a constant sum of units, such as points, among set of stimulus objects with respect to some criterion
4. **Q-Sort Technique** – sort object based on similarity with respect to some criterion

Non-Comparative Scales of Measurement

1. **Continuous Rating** – rate the objects by placing a mark at the appropriate position on a continuous line that runs from one extreme of the criterion variable to the other
- e.g., Rating Guardians of the Galaxy as the best Marvel Movie of Phase 4

2. **Itemized Rating** – having numbers or brief descriptions associated with each category
- e.g., 1 if your like the item the most, 2 if so-so, 3 if you hate it

3. **Likert Scale** – indicate their own attitudes by checking how strongly they agree or disagree with carefully worded statements that range from very positive to very negative towards attitudinal object
- principle of measuring attitudes by asking people to respond to a series of statements about a topic, in terms of the extent to which they agree with them

4. **Visual Analogue Scale** – a 100-mm line that allows subjects to express the magnitude of an experience or belief
5. **Semantic Differential Scale** – derive respondent's attitude towards the given object by asking him to select an appropriate position on a scale between two bipolar opposites
6. **Staple Scale** – developed to measure the direction and intensity of an attitude simultaneously
7. **Summative Scale** – final score is obtained by summing the ratings across all the items
8. **Thurstone Scale** – involves the collection of a variety of different statements about a phenomenon which are ranked by an expert panel in order to develop the questionnaire
- allows multiple answers

9. **Ipsative Scale** – the respondent must choose between two or more equally socially acceptable options

III. Test Tryout - the test should be tried out on people who are similar in critical respects to the people for whom the test was designed
- An informal rule of thumb should be no fewer than 5 and preferably as many as 10 for each item (the more, the better)

- Risk of using few subjects = phantom factors emerge
- Should be executed under conditions as identical as possible
- A good test item is one that answered correctly by high scorers as a whole
- *Empirical Criterion Keying:* administering a large pool of test items to a sample of individuals who are known to differ on the construct being measured
- *Item Analysis:* statistical procedure used to analyze items, evaluate test items
- *Discriminability Analysis:* employed to examine correlation between each item and the total score of the test
- *Item:* suggest a sample of behavior of an individual
- *Table of Specification:* a blueprint of the test in terms of number of items per difficulty, topic importance, or taxonomy
- **Guidelines for Item writing:** Define clearly what to measure, generate item pool, avoid long items, keep the level of reading difficulty appropriate for those who will complete the test, avoid double-barreled items, consider making positive and negative worded items
- *Double-Barreled Items:* items that convey more than one ideas at the same time
- *Item Difficulty:* defined by the number of people who get a particular item correct
- *Item-Difficulty Index:* calculating the proportion of the total number of testtakers who answered the item correctly; The larger, the easier the item
- *Item-Endorsement Index* for personality testing, percentage of individual who endorsed an item in a personality test
- The optimal average item difficulty is approx. 50% with items on the testing ranging in difficulty from about 30% to 80%

Item Difficulty Range	Level of Difficulty
0.0-0.19	Very difficult
0.20-0.39	Difficult
0.40-0.60	Average/moderately difficult
0.61-0.79	Easy
0.80-1.0	Very easy

- *Omnibus Spiral Format:* items in an ability are arranged into increasing difficulty
- *Item-Reliability Index:* provides an indication of the internal consistency of a test
- The higher Item-Reliability index, the greater the test's internal consistency
- *Item-Validity Index:* designed to provide an indication of the degree to which a test is measure what it purports to measure
- The higher Item-Validity index, the greater the test's criterion-related validity
- *Item-Discrimination Index:* measure of item discrimination; measure of the difference between the proportion of high scorers answering an item correctly and the proportion of low scorers answering the item correctly
- *Extreme Group Method:* compares people who have done well with those who have done poorly
- *Discrimination Index:* difference between these proportion
- *Point-Biserial Method:* correlation between a dichotomous variable and continuous variable

Range	Verbal Description
.40 & above	Very Good Item
.30 - .39	Good Item
.20 - .29	Fair Item
.09 - .19	Poor Item

- *Item-Characteristic Curve:* graphic representation of item difficulty and discrimination
- *Guessing:* one that eluded any universally accepted solutions
- Item analyses taken under speed conditions yield misleading or uninterpretable results
- Restrict item analysis on a speed test only to the items completed by the testtaker
- Test developer ideally should administer the test to be item-analyzed with generous time limits to complete the test
- *Optimal Difficulty Index Formula: [(1+chance of getting it correct)/2]*
  - Multiple choice: 0.625
  - True or False (Binary): 0.75
  - Three choices: 0.665
  - Five choices: 0.6

Scoring Items/Scoring Models

1. **Cumulative Model –** testtaker obtains a measure of the level of the trait; thus, high scorers may suggest high level in the trait being measured
2. **Class Scoring/Category Scoring –** testtaker response earn credit toward placement in a particular class or category with other testtaker whose pattern of responses is similar in some way

3. **Ipsative Scoring** – compares testtaker's score on one scale within a test to another scale within that same test, two unrelated constructs

IV. Test Revision – characterize each item according to its strength and weaknesses
- As revision proceeds, the advantage of writing a large item pool becomes more apparent because some items were removed and must be replaced by the items in the item pool
- Administer the revised test under standardized conditions to a second appropriate sample of examinee
- *Cross-Validation:* revalidation of a test on a sample of testtakers other than those on who test performance was originally found to be a valid predictor of some criterion; often results to validity shrinkage
- *Validity Shrinkage:* decrease in item validities that inevitably occurs after cross-validation
- *Co-validation:* conducted on two or more test using the same sample of testtakers
- *Co-norming:* creation of norms or the revision of existing norms
- *Anchor Protocol:* test protocol scored by highly authoritative scorer that is designed as a model for scoring and a mechanism for resolving scoring discrepancies
- *Scoring Drift:* discrepancy between scoring in an anchor protocol and the scoring of another protocol
- *Differential Item Functioning:* item functions differently in one group of testtakers known to have the same level of the underlying trait
- *DIF Analysis:* test developers scrutinize group by group item response curves looking for DIF Items
- *DIF Items:* items that respondents from different groups at the same level of underlying trait have different probabilities of endorsing a function of their group membership

- **Computerized Adaptive Testing** – refers to an interactive, computer administered test-taking process wherein items presented to the testtaker are based in part on the testtaker's performance on previous items
  - The test administered may be different for each testtaker, depending on the test performance on the items presented
  - Reduces floor and ceiling effects
  - *Floor Effects:* occurs when there is some lower limit on a survey or questionnaire and a large percentage of respondents score near this lower limit (testtakers have low scores)
  - *Ceiling Effects:* occurs when there is some upper limit on a survey or questionnaire and a large percentage of respondents score near this upper limit (testtakers have high scores)
  - *Item Branching:* ability of the computer to tailor the content and order of presentation of items on the basis of responses to previous items
  - *Routing Test:* subtest used to direct or route the testtaker to a suitable level of items
  - *Item-Mapping Method:* setting cut scores that entails a histographic representation of items and expert judgments regarding item effectiveness
- **Basal Level** – the level of which a the minimum criterion number of correct responses is obtained
- **Computer Assisted Psychological Assessment** – standardized test administration is assured for testtakers and variation is kept to a minimum
  - Test content and length is tailored according to the taker's ability

Statistics

- **Measurement** – the act of assigning numbers or symbols to characteristics of things according to rules

Descriptive Statistics – methods used to provide concise description of a collection of quantitative information

Inferential Statistics – method used to make inferences from observations of a small group of people known as sample to a larger group of individuals known as population

- **Magnitude** – the property of "moreness"
- **Equal Intervals** – the difference between two points at any place on the scale has the same meaning as the difference between two other points that differ by the same number of scale units
- **Absolute 0** – when nothing of the property being measured exists
- **Scale** – a set of numbers who properties model empirical properties of the objects to which the numbers are assigned

Continuous Scale – takes on any value within the range and the possible value within that range is infinite
- used to measure a variable which can theoretically be divided

Discrete Scale – can be counted; has distinct, countable values
- used to measure a variable which cannot be theoretically be divided

- **Error –** refers to the collective influence of all the factors on a test score or measurement beyond those specifically measured by the test or measurement
  - Degree to which the test score/measurement may be wrong, considering other factors like state of the testtaker, venue, test itself etc.
  - Measurement with continuous scale always involve with error

Four Levels of Scales of Measurement
Nominal – involve classification or categorization based on one or more distinguishing characteristics - Label and categorize observations but do not make any quantitative distinctions between observations - mode
Ordinal - rank ordering on some characteristics is also permissible - median
Interval - contains equal intervals, has no absolute zero point (even negative values have interpretation to it) - Zero value does not mean it represents none
Ratio - has true zero point (if the score is zero, it means none/null) - Easiest to manipulate

- **Distribution –** defined as a set of test scores arrayed for recording or study
- **Raw Scores –** straightforward, unmodified accounting of performance that is usually numerical
- **Frequency Distribution –** all scores are listed alongside the number of times each score occurred
- **Independent Variable –** being manipulated in the study
- **Quasi-Independent Variable –** nonmanipulated variable to designate groups
  - *Factor***:** for ANOVA

Post-Hoc Tests – used in ANOVA to determine which mean differences are significantly different
Tukey's HSD test – allows the compute a single value that determines the minimum difference between treatment means that is necessary for significance

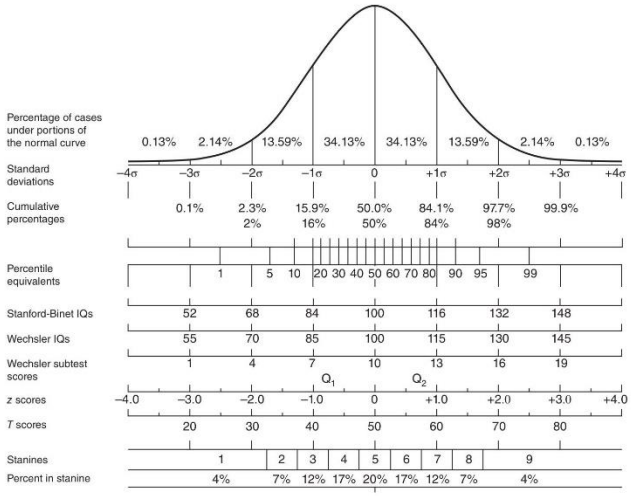
- **Measures of Central Tendency –** statistics that indicates the average or midmost score between the extreme scores in a distribution
  - Goal: Identify the most typical or representative of entire group

Mean – the average of all the raw scores - Equal to the sum of the observations divided by the number of observations - Interval and ratio data (when normal distribution) - Point of least squares - Balance point for the distribution

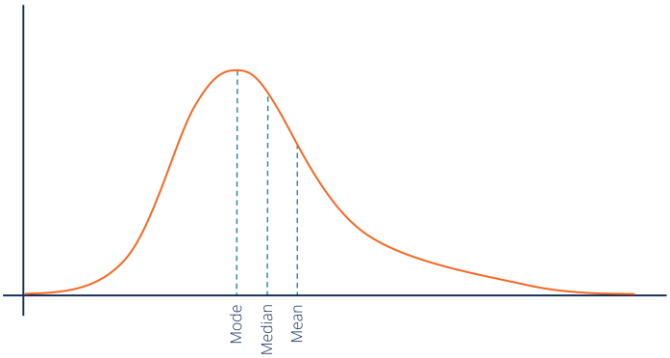
Median – the middle score of the distribution - Ordinal, Interval, Ratio - Useful in cases where relatively few scores fall at the high end of the distribution or relatively few scores fall at the low end of the distribution - In other words, for extreme scores, use median (skewed) - Identical for sample and population - Also used when there has an unknown or undetermined score - Used in "open-ended" categories (e.g., 5 or more, more than 8, at least 10) - For ordinal data
Mode – most frequently occurring score in the distribution - *Bimodal Distribution:* if there are two scores that occur with highest frequency - Not commonly used - Useful in analyses of qualitative or verbal nature - For nominal scales, discrete variables - Value of the mode gives an indication of the shape of the distribution as well as a measure of central tendency

- **Variability –** an indication how scores in a distribution are scattered or dispersed
- **Measures of Variability –** statistics that describe the amount of variation in a distribution
- **Range –** equal to the difference between highest and the lowest score
  - Provides a quick but gross description of the spread of scores
  - When its value is based on extreme scores of the distribution, the resulting description of variation may be understated or overstated
- **Quartile –** dividing points between the four quarters in the distribution
  - Specific point
  - *Quarter:* refers to an interval
  - *Interquartile Range:* measure of variability equal to the difference between Q3 and Q1
  - *Semi-interquartile Range:* equal to the interquartile range divided by 2
- **Standard Deviation –** equal to the square root of the average squared deviations about the mean
  - Equal to the square root of the variance
  - *Variance:* equal to the arithmetic mean of the squares of the differences between the scores in a distribution and their mean
  - Distance from the mean
- **Normal Curve –** also known as Gaussian Curve
- Bell-shaped, smooth, mathematically defined curve that is highest at its center

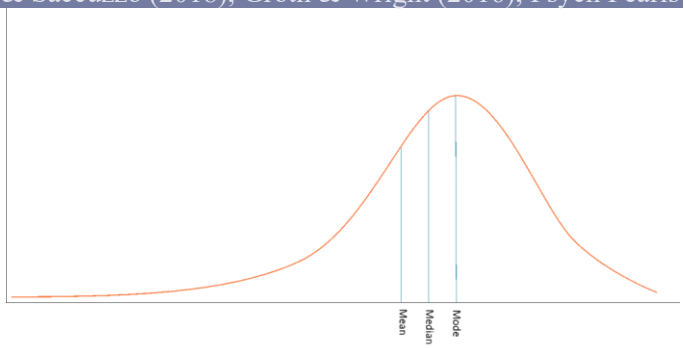
- Asymptotically = approaches but never touches the axis
- **Tail –** 2 – 3 standard deviations above and below the mean



- **Symmetrical Distribution –** right side of the graph is mirror image of the left side
  - Has only one mode and it is in the center of the distribution
  - Mean = median = mode
- **Skewness –** nature and extent to which symmetry is absent
- **Positive Skewed –** few scores fall the high end of the distribution
  - The exam is difficult
  - More items that was easier would have been desirable in order to better discriminate at the lower end of the distribution of test scores

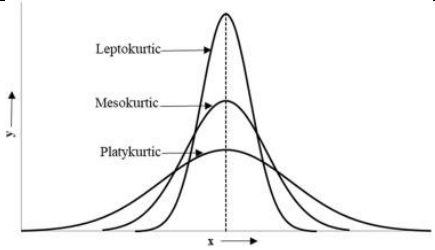


  - Mean > Median > Mode
- **Negative Skewed –** when relatively few of the scores fall at the low end of the distribution
  - The exam is easy
  - More items of a higher level of difficulty would make it possible to better discriminate between scores at the upper end of the distribution



  - Mean < Median < Mode
- Skewed is associated with abnormal, perhaps because the skewed distribution deviates from the symmetrical or so-called normal distribution
- **Kurtosis –** steepness if a distribution in its center

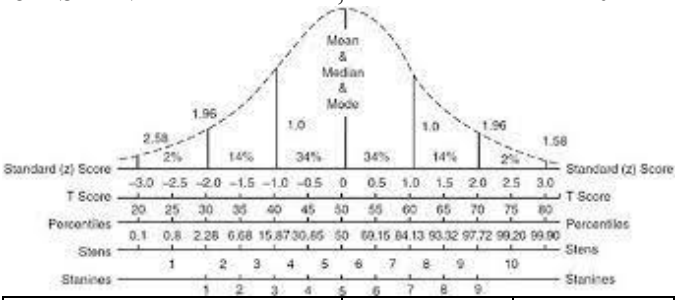
Platykurtic – relatively flat
Leptokurtic – relatively peaked
Mesokurtic – somewhere in the middle



  - High Kurtosis = high peak and fatter tails
  - Lower Kurtosis = rounded peak and thinner tails
- **Standard Score –** raw score that has been converted from one scale to another scale
- **Z-Scores –** results from the conversion of a raw score into a number indicating how many SD units the raw score is below or above the mean of the distribution
  - Identify and describe the exact location of each score in a distribution
  - Standardize an entire distribution
  - Zero plus or minus one scale
  - Have negative values
  - Requires that we know the value of the variance to compute the standard error
- **T-Scores –** a scale with a mean set at 50 and a standard deviation set at 10
  - Fifty plus or minus 10 scale
  - 5 standard deviations below the mean would be equal to a t-score of 0
  - Raw score that fell in the mean has T of 50
  - Raw score 5 standard deviations about the mean would be equal to a T of 100
  - No negative values
  - Used when the population or variance is unknown

Source: Cohen & Swerdlik (2018), Kaplan & Saccuzzo (2018), Groth & Wright (2016), Psych Pearls

- **Stanine** – a method of scaling test scores on a nine-point standard scale with a mean of five (5) and a standard deviation of two (2)
- **Linear Transformation** – one that retains a direct numerical relationship to the original raw score
- **Nonlinear Transformation** – required when the data under consideration are not normally distributed
- Normalizing the distribution involves stretching the skewed curve into the shape of a normal curve and creating a corresponding scale of standard scores, a scale that is technically referred to as **Normalized Standard Score Scale**
- Generally preferrable to fine-tune the test according to difficulty or other relevant variables so that the resulting distribution will approximate the normal curve
- **STEN** – standard to ten; divides a scale into 10 units

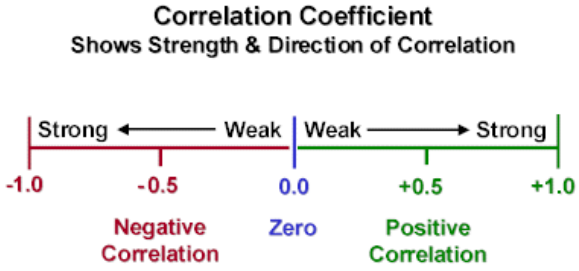


	Mean	SD
Z-Score	**0**	**1**
T-Score	**50**	**10**
Stanine	**5**	**2**
STEN	**5.5**	**2**
IQ	**100**	**15**
GRE or SAT	**500**	**100**

- **Hypothesis Testing** – statistical method that uses a sample data to evaluate a hypothesis about a population

Alternative Hypothesis – states there is a change, difference, or relationships
Null Hypothesis – no change, no difference, or no relationship

- **Alpha Level or Level of Significance** – used to define concept of "very unlikely" in a hypothesis test
- **Critical Region** – composed of extreme values that are very unlikely to be obtained if the null hypothesis is true
- If sample data fall in the critical region, the null hypothesis is rejected
- The alpha level for a hypothesis test is the probability that the test will lead to a Type I error
- **Directional Hypothesis Test or One-Tailed Test** – statistical hypotheses specify either an increase or a decrease in the population mean
- **T-Test** – used to test hypotheses about an unknown population mean and variance
  - Can be used in "before and after" type of research
  - Sample must consist of independent observations—that is, if there is not consistent, predictable relationship between the first observation and the second
  - The population that is sampled must be normal
  - If not normal distribution, use a large sample
- **Correlation Coefficient** – number that provides us with an index of the strength of the relationship between two things
- **Correlation** – an expression of the degree and direction of correspondence between two things
  - + & - = direction
  - Number anywhere to -1 to 1 = magnitude
  - **Positive** – same direction, either both going up or both going down
  - **Negative** – Inverse Direction, either DV is up and IV goes down or IV goes up and DV goes down
  - 0 = no correlation



- **Pearson r/Pearson Correlation Coefficient/Pearson Product-Moment Coefficient of Correlation** – used when two variables being correlated are continuous and linear
  - Devised by Karl Pearson
  - Coefficient of Determination – an indication of how much variance is shared by the X- and Y-variables
- **Spearman Rho/Rank-Order Correlation Coefficient/Rank-Difference Correlation Coefficient** – frequently used if the sample size is small and when both sets of measurement are in ordinal
  - Developed by Charles Spearman
- **Outlier** – extremely atypical point located at a relatively long distance from the rest of the coordinate points in a scatterplot

- **Regression Analysis** – used for prediction
  - Predict the values of a dependent or response variable based on values of at least one independent or explanatory variable
  - *Residual:* the difference between an observed value of the response variable and the value of the response variable predicted from the regression line
  - The Principle of Least Squares
  - *Standard Error of Estimate:* standard deviation of the residuals in regression analysis
  - *Slope:* determines how much the Y variable changes when X is increased by 1 point
- **T-Test (Independent)** – comparison or determining differences
  - 2 different groups/independent samples + interval/ratio scales (continuous variables)

Equal Variance – 2 groups are equal

Unequal Variance – groups are unequal

- **T-test (Dependent)/Paired Test** – one groups nominal (either matched or repeated measures) + 2 treatments
- **One-Way ANOVA** – 3 or more IV, 1 DV comparison of differences
- **Two-Way ANOVA** – 2 IV, 1 DV
- **Critical Value** – reject the null and accept the alternative if [*obtained value* > *critical value*]
- **P-Value (Probability Value)** – reject null and accept alternative if [*p-value* < *alpha level*]
- **Norms** – refer to the performances by defined groups on a particular test

Percentiles – an expression of the percentage of people whose score on a tests or measure falls below a particular raw score

Age Norms – average performance of different samples of testtakers who were at various ages at the time the test was administered

Grade Norms – developed by administering the test to representative samples of children over a range of consecutive grade levels
- *Developmental Norms:* developed on the basis of any trait, ability, skill, or other characteristics that is presumed to develop, deteriorate, or otherwise be affected by chronical age, school grade, or stage of life

National norms – derived from a normative sample that was nationally representative of the population at the time the norming study was conducted

Subgroup Norms – normative sample can be segmented by any criteria initially used in selecting subjects for the sample

Local Norms – provide normative information with respect to the local population's performance on some tests

- **Age-Related Norms** – certain tests have different normative groups for age groups
- **Tracking** – tendency to stay at about the same level relative to one's peers

Norm-Referenced Tests – compares each person with the norm

Criterion-Referenced Tests – describes specific types of skills, tasks, or knowledge that the test taker can demonstrate

Selection of Assessment Methods and Tools and Uses, Benefits, and Limitations of Assessment tools and instruments (32)

Identify appropriate assessment methods, tools (2)

1. **Test** – measuring device or procedure
- *Psychological Test:* device or procedure designed to measure variables related to psychology

Ability or Maximal Performance Test – assess what a person can do

a. **Achievement Test** – measurement of the previous learning

b. **Aptitude** – refers to the potential for learning or acquiring a specific skill

c. **Intelligence** – refers to a person's general potential to solve problems, adapt to changing environments, abstract thinking, and profit from experience

Human Ability – considerable overlap of achievement, aptitude, and intelligence test

Typical Performance Test – measure usual or habitual thoughts, feelings, and behavior

Personality Test – measures individual dispositions and preferences

a. **Structured Personality tests** – provide statement, usually self-report, and require the subject to choose between two or more alternative responses

b. **Projective Personality Tests** – unstructured, and the stimulus or response are ambiguous

c. **Attitude Test** – elicit personal beliefs and opinions

d. **Interest Inventories** – measures likes and dislikes as well as one's personality orientation towards the world of work

- Purpose: for evaluation, drawing conclusions of some aspects of the behavior of a person, therapy, decision-making
- Settings: Industrial, Clinical, Educational, Counseling, Business, Courts, Research

Source: Cohen & Swerdlik (2018), Kaplan & Saccuzzo (2018), Groth & Wright (2016), Psych Pearls

- Population: Test Developers, Test Publishers, Test Reviewers, Test Users, Test Sponsors, Test Takers, Society

Levels of Tests

1. **Level A –** anyone under a direction of a supervisor or consultant
2. **Level B –** psychometricians and psychologists only
3. **Level C –** psychologists only

2. **Interview –** method of gathering information through direct communication involving reciprocal exchange
- can be structured, unstructured, semi-structured, or non-directive
- *Mental Status Examination:* determines the mental status of the patient
- *Intake Interview:* determine why the client came for assessment; chance to inform the client about the policies, fees, and process involved
- *Social Case:* biographical sketch of the client
- *Employment Interview:* determine whether the candidate is suitable for hiring
- *Panel Interview (Board Interview):* more than one interviewer participates in the assessment
- *Motivational Interview:* used by counselors and clinicians to gather information about some problematic behavior, while simultaneously attempting to address it therapeutically

Issues addressed by Structured Interviews:

1. *Criterion Variance:* amount of error caused by vague guidelines for exclusion or inclusion in different categories
2. *Information Variance:* variability in amount and type of information derived from interviews with clients

3. **Portfolio –** samples of one's ability and accomplishment
- Purpose: Usually in industrial settings for evaluation of future performance

4. **Case History Data –** refers to records, transcripts, and other accounts in written, pictorial, or other form that preserve archival information, official and informal accounts, and other data and items relevant to an assessee

5. **Behavioral Observation –** monitoring of actions of others or oneself by visual or electronic means while recording quantitative and/or qualitative information regarding those actions
- *Naturalistic Observation:* observe humans in natural setting

6. **Role Play –** defined as acting an improvised or partially improvised part in a simulated situation
- *Role Play Test:* assesses are directed to act as if they are in a particular situation
- Purpose: Assessment and Evaluation
- Settings: Industrial, Clinical
- Population: Job Applicants, Children

7. **Computers –** using technology to assess an client, thus, can serve as test administrators and very efficient test scorers

8. **Others:** videos, biofeedback devices

Intelligence Tests

Stanford-Binet Intelligence Scale 5th Ed. (SB-5)
[C]
- 2-85 years old
- individually administered
- norm-referenced
- age scale and point-scale format
- originally created to identify mentally disabled children in Paris
- 1908 Scale introduced Age Scale format and Mental Age
- 1916 scale significantly applied IQ concept
- Standard Scores: 100 (mean), 15 (SD)
- Scaled Scores: 10 (mean), 3 (SD)
- co-normed with Bender-Gestalt and Woodcock-Johnson Tests
- based on Cattell-Horn-Carroll Model of General Intellectual Ability
- no accommodations for PWDs
- 2 *routing tests* (task used to direct or route the examinee to a particular level of questions)
- w/ teaching items, floor level, and ceiling level
- provides behavioral observations during administration
- First ed lacks representativeness of the standardization sample
- first test to introduce the concept of an *alternate item*
- earlier versions employed *Ratio IQ*, which was based on the concept of *Mental Age*
- 1960 revision had consisted of single form and the use of *Deviation IQ*
- *Ratio IQ:* testtaker's mental age divided by chronical age, multiplied by 100
- *Deviation IQ:* reflects a comparison of the performance of the individual with the performance of others of the same age in the standardization sample
- SB-4 had *Point Scale*, which was the test was organized into subtests by category of item, not by age

at which most testtakers are presumed capable of responding in the way that is keyed as correct

SB5 Scales:
1. Verbal
2. Nonverbal
3. Full Scale IQ (FSIQ)

SB5 Factors:
1. Fluid Reasoning (FR)
2. Knowledge (KN)
3. Qualitative Reasoning (QR)
4. Visual-Spatial Processing (VS)
5. Working Memory (WM)

Measured IQ Range	Category
145–160	Very gifted or highly advanced
130–144	Gifted or very advanced
120–129	Superior
110–119	High average
90–109	Average
80–89	Low average
70–79	Borderline impaired or delayed
55–69	Mildly impaired or delayed
40–54	Moderately impaired or delayed

Wechsler Intelligence Scales (WAIS-IV, WPPSI-IV, WISC-V)
[C]
- WAIS (16-90 years old), WPPSI (2-6 years old), WISC (6-11)
- individually administered
- norm-referenced
- Standard Scores: 100 (mean), 15 (SD)
- Scaled Scores: 10 (mean), 3 (SD)
- addresses the weakness in Stanford-Binet
- could also assess functioning in people with brain injury
- evaluates patterns of brain dysfunction
- measure abilities that are likely to be lowered by brain damage
- useful in assessment of Alzheimer's Disease
- WAIS (1955), WAIS-R (1981), WAIS-III (1997), WAIS-IV (2008)
- the latest edition has updated norms, improved floor and ceilings, improved psychometric properties, reduced testing time, and co-normed with Weschler Individual Achievement Test-II)
- time-honored verbal versus performance IQ was eliminated (they do not purely measure but typically combine a number of different abilities)
- WISC (1949), WISC-R (1974), WISC-III (1991), WISC-IV (2003)

WAIS-IV Subtests:
1. Verbal Comprehension
2. Perceptual Reasoning
3. Working Memory
4. Processing Speed

WISC-IV Subtests:
1. Verbal Comprehension
2. Visual Spatial
3. Fluid Reasoning
4. Working Memory
5. Processing Speed

Table 5.5 WAIS-IV/WISC-V Intelligence Classifications

Classifications	More value-neutral terms	Corresponding IQ range
Very superior	Upper extreme	130+
Superior	Well above average	120–129
High average	High average	110–119
Average	Average	90–109
Low average	Low average	80–89
Borderline	Well below average	70–79
Extremely low	Lower extreme	69 and below

Source: Intelligence classifications adapted from WAIS-IV and WISC-V Record forms.

Successive Five-Level Interpretative Procedures:
Level I. Interpret the FSIQ
Level II. Interpret index scores and CHC groupings
Level III. Interpret subtest variability
Level IV. Qualitative/Process Analysis
Level V. Analysis Intrasubtest variability

Raven's Progressive Matrices (RPM)
[B]
- 4-90 years old
- CPM, 5-11 years old; SPM, 6-16 yrs old & 17 yrs old and older; APM, 12+ yrs old
- nonverbal test
- used to measure general intelligence & abstract reasoning
- multiple choice of abstract reasoning
- group test
- IRT-Based
- *Colored Progressive Matrices:* Used to assess the degree to which children and adults can think clearly, or the level to which their intellectual abilities have deteriorated
- *Standard Progressive Matrices:* more difficult that CPM
- *Advanced Progressive Matrices:* geared towards adults and teenagers of advanced intelligence

Culture Fair Intelligence Test (CFIT)
[B]

Source: Cohen & Swerdlik (2018), Kaplan & Saccuzzo (2018), Groth & Wright (2016), Psych Pearls

- Nonverbal instrument to measure your analytical and reasoning ability in the abstract and novel situations
- Measures individual intelligence in a manner designed to reduced, as much as possible, the influence of culture
- Individual or by group
- Aids in the identification of learning problems and helps in making more reliable and informed decisions in relation to the special education needs of children

Purdue Non-Language Test
[B]
- Joseph Tiffin, Alin Grubner & Kay Inaba
- 13 yrs old and above
- Designed to measure mental ability, since it consists entirely of geometric forms, assess the fluid intelligence
- mainly abstract reasoning
- determine the underlying logic of pattern and devise a solution
- 5 designs, two parallel forms, 48 items
- Culture-fair
- Self-Administering

Panukat ng Katalinuhang Pilipino (PKP)
- Aurora R. Palacio
- 16 yrs old and above
- Basis for screening, classifying, and identifying needs that will enhance the learning process
- In business, it is utilized as predictors of occupational achievement by gauging applicant's ability and fitness for a particular job
- Essential for determining one's capacity to handle the challenges associated with certain degree programs
- measures fluid, crystallized, and generalized intelligence

Subtests:
1. Vocabulary
2. Analogy
3. Numerical Ability
4. Nonverbal Ability

Wonderlic Personnel Test (WPT)
- Eldon Wonderlic
- Adults
- Assessing cognitive ability and problem-solving aptitude of prospective employees
- Multiple choice, answered in 12 minutes

Armed Services Vocational Aptitude Battery (ASVAB)
- Most widely used aptitude test in US
- Multiple-aptitude battery that measures developed abilities and helps predict future academic and occupational success in the military

Kaufman Assessment Battery for Children-II (KABC-II)
- Alan & Nadeen Kaufman
- 2 ½ yrs to 12 ½ yrs old
- for assessing cognitive development in children
- evaluate preschoolers, minority, and children with learning disabilities

Global Test Scores:
1. Sequential Processing Scales
2. Simultaneous Processing Scales
3. Achievement Scales
4. Mental Processing Composite

Differential Aptitude Scale – Fifth Edition (DAT-V)
[B]
- Harold G. Raven, John M. Oldfield, and John C. Raven
- 14 to 65 years old
- assess an individual's aptitudes across various domains, including verbal reasoning, numerical ability, abstract reasoning, mechanical reasoning, space relations, spelling, language usage, and clerical speed and accuracy
- provide insights into an individual's relative strengths and weaknesses in these areas to guide educational and career decisions
- norm-referenced

General Aptitude Test Battery (GATB)
- U.S. Employment Services
- measure of wide range of aptitudes and is used in such areas as occupational selection, rehabilitation, and vocational counseling

Aptitude Scores:
G – General Intelligence
V – Verbal Aptitude
N – Numerical Aptitude
S – Spatial Aptitude
P – Form Perception
Q – Clerical-Perception
K – Motor Coordination
F – Finger Dexterity
M – Manual Dexterity

California Verbal Learning Test (CVLT-II)
[C]
- 16 yrs old – 89 yrs old

Source: Cohen & Swerdlik (2018), Kaplan & Saccuzzo (2018), Groth & Wright (2016), Psych Pearls

- measures episodic verbal learning and memory and demonstrates sensitivity to a range of clinical conditions

Otis-Lennon Ability Test (OLSAT)
[C]
- Otis, A.S. & Lennon, R.T.
- designed to assess general mental ability or scholastic aptitude of pupils
- identify gifted children

Philippine Aptitude Classification Test (PACT)
- Ma. Lourdes M. Franco
- 14 yrs to 15 yrs old
- developed to measure student's abilities and help students decide on the course they will take after high school
- assumes that aptitudes are required in different combinations and in varying degrees for successful performance in different post-secondary courses

Flanagan Industrial Tests
- Flanagan, J.C.
- used for personnel selection programs, based on identified job elements

Job Elements
1. Arithmetic
2. Assembly
3. Components
4. Coordination
5. Electronics
6. Expression
7. Ingenuity
8. Inspection
9. Judgment and Comprehension
10. Mathematics and Reasoning
11. Mechanics
12. Memory Patterns
13. Planning
14. Precision
15. Scales
16. Tables
17. Vocabulary

Watson Glaser Critical Thinking Test (W-GCTA)
- Goodwin Watson & Edward Glaser
- 20 yrs old to 64 yrs old
- designed to assess a person's critical thinking abilities and is widely used across legal practices
- assesses ability for critical thinking, creating conclusions, analyzing strong and weak arguments, recognizing assumptions, and evaluating arguments

Thurstone Test of Mental Alertness (TMA)
- Louis Thurstone
- Adults

Four Job-Related Tasks assessed by TMA Test:
1. Adjusting to new situations
2. Learning new skills quickly
3. Understanding complex or subtle relationships
4. Thinking flexibly

Personality Tests

Minnesota Multiphasic Personality Inventory (MMPI-2)
[C]
- Starke Hathaway and J. Charnley McKinley
- 16 years old and older
- Multiphasic personality inventory intended for used with both clinical and normal populations to identify sources of maladjustment and personal strengths
- Help in diagnosing mental health disorders, distinguishing normal from abnormal
- elicits a wide range of self-descriptions scored to give a quantitative measurement of an individual's level of emotional adjustment and attitude toward test-taking
- should be administered to someone with no guilt feelings for creating a crime
- individual or by groups
- Original MMPI had 13 scales, whilst the latest ver (MMPI-II) maintained the original 10 scales
- MMPI (1939),
- most important approach taken during construction of the MMPI was *Empirical Criterion Keying* (development, selection, and scoring of items within the scales was based on some external criterion of reference)

Clinical Scales:
1. *Hypochondriasis* (Hs) – present multiple, vague, and chronic physical problems
2. *Depression* (D) – depressed mood, low self-esteem, lethargy, and feelings of guilt
3. *Hysteria* (Hy) – develop physical symptoms in reaction to stress and can be dependent, naïve, infantile, and narcissistic
4. *Psychopathic Deviate* (Pd) – associated with antisocial behavior
5. *Masculinity/Femininity* (Mf) – interests more traditionally viewed as feminine or masculine
6. *Paranoia* (Pa) – suspicious, aloof, guarded, and overly sensitive

7. *Psychasthenia* (Anxiety, Depression, OCD) (Pt) – tense, anxious, ruminative, obsessive, phobic, and rigid
8. *Schizophrenia* (Sc) – withdrawn, moody, and confused
9. *Hypomania or Mania* (Ma) – sociable and optimistic, though can be manipulative and grandiose
10. *Social Introversion* (Si) – introverted, withdrawn, submissive, over-controlled, tense, inflexible (high scorers)

Validity/Dissimulation scales: understand how genuine a test taker's answers are
1. *Lie Scale (L Scale):* items that are somewhat negative but apply to most people; assess the likelihood of the test taker to approach the instrument with defensive mindset
- High in L scale = faking good
- Low in L Scale = frank and open regarding responses to items, can be sarcastic and cynical
2. *Infrequency Scale (F Scale):* reveal inconsistencies in answer patterns and could also indicate severe distress or psychopathology
- measures the extent to which a person answers in an atypical and deviant manner
- High in F scale = faking bad, severe distress or psychopathology
- Moderate Scores = draw attention to distress as cry for help, may be rebellious, antisocial, curious, psychologically sophisticated
- Low Scores = perceive the world as most other people do, possible denial of difficulties
3. *Superlative Self Presentation Scale (S Scale):* a measure of defensiveness; Superlative Self-Presentation to see if you intentionally distort answers to look better (Social Desirability)
4. *Correction Scale (K Scale):* reflection of the frankness of the testtaker's self-report
- K Scale = reveals a person's defensiveness around certain questions and traits; also faking good
- K scale sometimes used to correct scores on five clinical scales. The scores are statistically corrected for an individual's overwillingness or unwillingness to admit deviance
5. *"Cannot Say" (CNS or ?) Scale:* measures how a person doesn't answer a test item; number of items left unanswered
- client might have difficulties with reading, psychomotor retardation, or extreme defensiveness
6. *True Response Inconsistency (TRIN):* five true, then five false answers
- Very High = person is indiscriminately answering "true" to the items
7. *Variable Response Inconsistency (VRIN or VRIN):* random true or false
- expected to be answered in a consistent manner if the person is approaching the testing in a valid manner
- if a person answers in the opposite direction, then it indicates an inconsistent response and is, therefore, scored as 1 raw score point on the VRIN Scale
- High VRIN = indiscriminate responding, profile should be considered invalid and should not be interpreted
8. *Infrequency-Psychopathology Scale (Fp):* reveal intentional or unintentional over-reporting
- High Fp = high probability of faking or exaggerating psychopathology, even among psychiatric patients
9. *Fake Bad Scale (FBS):* "symptom validity scale" designed to detect intentional over-reporting of symptoms
- detect personal injury claimants who were exaggerating their difficulties
10. *Back Page Infrequency (Fb):* reflects significant change in the testtaker's approach to the latter part of the test
- identify a "fake bad" mode of responding for the last 197 items
- High Fb = exaggeration of psychopathology

Higher-Order Scales
1. Emotional/Internalizing Dysfunction (EID)
2. Thought Dysfunction (THD)
3. Behavioral/Externalizing Dysfunction (BXD)

Personality Assessment Inventory (PAI)
- Leslie Morey
- 18 to 89, PAI-A for 12-18 years old
- self-administered, paper-and-pencil/online test composed of 344 statements for which the respondent must choose who true each is for him or her
- assesses psychopathological syndromes and provide information relevant for clinical diagnosis, treatment planning, and screening for psychopathology
- items for each scale were developed based on extensive reviews of both historical, conceptual literature and contemporary, empirical literature, focusing on the concepts that are central and core to the concepts of each construct

Validity Scales:
1. Inconsistency (ICN)
2. Infrequency (INF)
3. Negative Impression (NIM)

4. Positive Impression (PIM)

Clinical Scales:
1. Somatic Complaints (SOM)
2. Anxiety (ANX)
3. Anxiety-Related Disorders (ARD)
4. Depression (DEP)
5. Mania (MAN)
6. Paranoia (PAI)
7. Schizophrenia (SCZ)
8. Borderline Features (BOR)
9. Antisocial Features (ANT)
10. Alcohol Problems (ALC)
11. Drug Problems (DRG)

Treatment Scales:
1. Aggression (AGG)
2. Suicidal Ideation (SUI)
3. Stress (STR)
4. Nonsupport (NON)
5, Treatment Rejection (RXR)

Interpersonal Scales:
1. Dominance (DOM)
2. Warmth (WRM)

Myers-Briggs Type Indicator (MBTI)
- Katherine Cook Briggs and Isabel Briggs Myers
- Self-report inventory designed to identify a person's personality type, strengths, and preferences
- *Extraversion-Introversion Scale:* where you prefer to focus your attention and energy, the outer world and external events or your inner world of ideas and experiences
- *Sensing-Intuition Scale:* how do you take inform, you take in or focus on interpreting and adding meaning on the information
- *Thinking-Feeling Scale:* how do you make decisions, logical or following what your heart says
- *Judging-Perceiving Scale:* how do you orient the outer world? What is your style in dealing with the outer world – get things decided or stay open to new info and options?

Edward's Preference Personality Schedule (EPPS) [B]
- Adults
- Edwards, A. L.
- designed primarily as an instrument for research and counselling purposes to provide quick and convenient measures of a number of relatively normal personality variables
- based of Murray's Need Theory
- Objective, forced-choice inventory for assessing the relative importance that an individual places on 15 personality variables
- Useful in personal counselling and with non-clinical adults
- Individual

Guilford-Zimmerman Temperament Survey (GZTS)
- J.P. Guilford & Wayne Zimmerman
- 16 yrs and older
- items are stated affirmatively rather than in question form, using the 2nd person pronoun
- factor analysis

Personality Traits measured:
1. General Activity (G)
2. Restraint (R)
3. Ascendance (A)
4. Sociability (S)
5. Emotional Stability (E)
6. Objectivity (O)
7. Friendliness (F)
8. Thoughtfulness (T)
9. Personal Relations (P)
10. Masculinity (M)

NEO Personality Inventory (NEO-PI-R)
- Costa and McCrae
- 17 to 89 years old, NEO-PI-3 12 to 99 years old
- Standard questionnaire measure of the Five Factor Model, provides systematic assessment of emotional, interpersonal, experiential, attitudinal, and motivational styles
- gold standard for personality assessment
- Self-Administered
- Brief Ver: NEO-FFI
- T-scores

Neuroticism:
- identifies individuals who are prone to psychological distress
- measures the tendency toward emotional instability, turmoil, and general distress
- high = anxious, emotionally labile, quick to anger, sad, impulsive, low self-worth
- low = emotionally stable and secure, less prone to sadness
- extremely low = less productive, no sense of urgency

1. *Anxiety (N1):* measure proneness to becoming tense, jittery, nervous, fearful
2. *Angry Hostility (N2):* tendency toward anger bitterness, and resentment

Source: Cohen & Swerdlik (2018), Kaplan & Saccuzzo (2018), Groth & Wright (2016), Psych Pearls

3. *Depression (N3):* likelihood to experience the range of depressive affects, including sadness, loneliness, hopelessness, helplessness, worthlessness, etc.
4. *Self-Consciousness (N4):* measures discomfort with social awkwardness
5. *Impulsiveness (N5):* measures the degree to which individuals have difficulty resisting their urges
6. *Vulnerability (N6):* measures the degree to which individuals feel capable or not of coping with stress

Extraversion:
- quantity and intensity of energy directed
- measures the degree to which individuals are not only outgoing and sociable, but also assertive, upbeat, warm, and friendly
- high = talkative, warm, and friendly, leaders, but can be quite socially dominant and aggressive
- low = prefers to be on their own, reserved
1. *Warmth (E1):* measures comfort with interpersonal intimacy and closeness
2. *Gregariousness (E2):* measures the preference for having other people around
3. *Assertiveness (E3):* tendency to make oneself heard and known in social situations
4. *Activity (E4):* measures the amount of energy and gusto with which individuals live their lives
5. *Excitement-Seeking (E5):* measures the need for and enjoyment in high-stimulating activities
6. *Positive Emotions (E6):* tendency to experience positive emotions like happiness, joy, bliss, and love

Openness to Experience:
- active seeking and appreciation of experiences for their own sake
- encompasses many different personality traits that have been researched in the literature, including imagination, curiosity, attunement toward personal emotions, and preference for abstract thinking
- high = intellectually and creatively curious, open to new ideas and to values and theories that may contradict or challenge their own
- low = conservative and conventional, realistic and level-headed in solving problems
1. *Fantasy (O1):* how active one's imagination is and how much fantasy is used not as an escape from reality, but as a way to create, solve problems, and even interact with the world
2. *Aesthetics (O2):* measures the interest in and absorption by art, beauty, and even the inherent beauty in nature
3. *Feelings (O3):* measures openness to one's inner emotional life, including both allowing oneself to feel deep emotions and valuing them as integral to the entire, full human experience
4. *Actions (O4):* measures the behavioral aspects of openness, relating to trying novel and unknown activities, foods, and places
5. *Ideas (O5):* measure a cognitive aspect of openness, related to intellectual curiosity
6. *Values (O6):* measure the willingness to re-examine one's values of people with different worldviews, different cultures, and a changing world, in general

Agreeableness:
- the kind of interactions an individual prefers from compassion to tough mindedness
- measures both attitudes about the trustworthiness and general goodness of others and behaviors related to respecting, empathizing with, and deferring to others
- high = sympathetic, willing to help, cooperative, believes are others are generally decent and hones
- low = skeptical of others, expecting competition and challenge from people around them, sarcastic and stubborn
1. *Trust (A1):* faith in the goodness of the human spirit
2. *Straightforwardness (A2):* measures directness, honesty, and genuineness
3. *Altruism (A3):* measure genuine concern for the well-being of others
4. *Compliance (A4):* related to the way individuals react to conflict with others
5. *Modesty (A5):* measures the outward-facing trait of humility, including beliefs about being better than or equal to others and whatever they would be likely to boast about their achievements
6. *Tender-Mindedness (A6):* measure sympathy and concern for others

Conscientiousness:
- degree of organization, persistence, control, and motivation in goal-directed behavior
- measures an array of traits related to both an orientation toward accomplishing things and the behavioral correlates of doing so successfully
- high = motivated to achieve their goals, plans and organize well, extremely reliable but can also be moralistic and judgmental of both themselves and others

- low = often unprepared and disorganized, do not think through potential consequences before acting, not overly driven to succeed
1. *Competence (C1):* measure the feeling that one is generally effective and capable to succeed in tasks and life in general
2. *Order (C2):* measures preference for neatness, tidiness, and orderliness
3. *Dutifulness (C3):* measures the degree to which individuals do things that they feel they should do
4. *Achievement Striving (C4):* relates to an attitude of aspiration and striving to succeed in their goals
5. *Self-Discipline (C5):* ability to actually follow through on a plan once it is set
6. *Deliberation (C6):* measures the degree to which individuals think and plan out carefully before acting

Panukat ng Ugali at Pagkatao/Panukat ng Pagkataong Pilipino
- 13 years and older
- Indigenous personality test
- Tap specific values, traits and behavioral dimensions related or meaningful to the study of Filipinos

1. Pagkamaalalahanin (Thoughtfulness)
2. Pagkamaayos (Organized)
3. Pagkamadaldal (Social Curiosity)
4. Pagkamagalang (Respectfulness)
5. Pagkamahinahon (Emotional Stability)
6. Pagkamalikhain (Creativity)
7. Pagkamapagkumbaba (Humility)
8. Pagkamapagsapalaran (Risk-Taking)
9. Pagkamadamdamin (Sensitiveness)
10. Pagkamasiyahin (Cheerfulness)
11. Pagkamasikap (Achievement-Oriented)
12. Pagkamasunurin (Obedience)
13. Pagkamatalino (Intelligence)
14. Pagkamatapat (Honesty)
15. Pagkamatiyaga (Patience)
16. Pagkamatulungin (Helpfulness)
17. Pagkamaunawain (Capacity for Understanding)
18. Pagkapalakaibigan (Sociability)
19. Pagkaresponsable (Responsibility)

Sixteen Personality Factor Questionnaire (16PF)
- Raymond Cattell
- constructed through factor analysis
- Evaluates a personality on two levels of traits

Primary Scales:
A Warmth (where)
B Reasoning (r)
C Emotional Stability (everyone?)
E Dominance (does)
F Liveliness (life)
G Rule-Consciousness (really)
H Social Boldness (suck?)
I Sensitivity (satisfy my)
L Vigilance (virgin)
M Abstractedness (ass)
N Privateness (putangina)
0 Apprehension (ay wow)
Q1 Openness to change (oh)
Q2 Self-Reliance (shit!)
Q3 Perfectionism (perfect)
Q4 Tension (timing!)

Global Scales:
1. Extraversion
2. Anxiety
3. Tough-Mindedness
4. Independence
5. Self-Control

Response Style Indices:
a. *Impression Management* – answering in socially desirable or undesirable way
- sten score of 9 or 10 at or above the 95th percentile denotes *Social Desirability*
- a sten score of 1 or 2, or a score below the 5th percentile means that the client has attempted to give a *socially unfavorable* image
b. *Acquiescence* – agree to questions regardless of the content
- scores at or above 95th percentile
- count each "true"
- confused when completing the questionnaire
c. *Infrequency* – detects unusual responses, responded randomly or indecisively
- count "B or ?" response
- scores at or above 95th percentile
- poor comprehension
- trying to avoid making a wrong impression

Big Five Inventory-II (BFI-2)
- Soto & John
- Assesses big 5 domains and 15 facets
- for commercial purposes to researches and students

California Psychological Inventory (CPI-434) [C]
- Harrison G. Gough
- 13 years old and older
- create efficient and productive organizations, promote teamwork, build leadership competencies,

and find and develop employees who are destined for success
- CPI-260, short form, 40-60 minutes
- T-Scores

Folk Scales:
1. Capacity for Status
2. Sociability
3. Social Presence
4. Self-Acceptance
5. Sense of Well-Being
6. Responsibility
7. Socialization
8. Self-Control
9. Tolerance
10. Good Impression
11. Communality
12. Achievement via Conformance
13. Achievement via Independence
14. Intellectual Efficiency
15. Psychological Mindedness
16. Femininity/Masculinity Independence
17. Flexibility and Empathy

Strong Campbell Interest Inventory (SCII)
- E.K. Strong
- 15 years old and older
- help individuals in identifying their work personality by exploring their interests in six broad areas:
RIASEC

Thurstone Interest Schedule (TIS)
[A]
- Thurstone, L.L.
- checklist by which a person can systematically clarify his understanding of his vocational interest
- designed as a counseling instrument to be used in situations in which the client-counselor relationship is such that straightforward and honest

Basic Personality Inventory (BPI)
[C]
- Adults & Adolescents
- Douglas Jackson
- self-report measure of general domain of psychopathology
- 240 true/false items, 11 substantive clinical scales and one critical item scale

Dimensions:
1. Alienation
2. Anxiety
3. Denial
4. Depression
5. Deviation
6. Hypochondriasis
7. Impulse Expression
8. Interpersonal Problems
9. Persecutory Ideas
10. Self-Depreciation
11. Social Introversion
12. Thinking Disorder

Projective Tests

Rorshcach Inkblot Test
[C]
- Hermann Rorschach
- 5 years and older
- subjects look at 10 ambiguous inkblot images and describe what they see in each one
- once used to diagnose mental illnesses like schizophrenia
- *Exner System:* coding system used in this test
- *Content:* the name or class of objects used in the patient's responses

Content:
1. Nature
2. Animal Feature
3. Whole Human
4. Human Feature
5. Fictional/Mythical Human Detail
6. Sex

Determinants:
1. Form
2. Movement
3. Color
4. Shading
5. Pairs and Reflections

Location:
1. W – the whole inkblot was used to depict an image
2. D – commonly described part of the blot was used
3. Dd – an uncommonly described or unusual detail was used
4. S – the white space in the background was used

Thematic Apperception Test
[C]
- Christiana Morgan and Henry Murray
- 5 and above
- 31 picture cards serve as stimuli for stories and descriptions about relationships or social situations
- popularly known as the picture interpretation technique because it uses a standard series of

Source: Cohen & Swerdlik (2018), Kaplan & Saccuzzo (2018), Groth & Wright (2016), Psych Pearls

provocative yet ambiguous pictures about which the subject is asked to tell a story
- also modified African American testtakers

Children's Apperception Test
- Bellak & Bellak
- 3-10 years old
- based on the idea that animals engaged in various activities were useful in stimulating projective storytelling by children

Hand Test
- Edward Wagner
- 5 years old and above
- used to measure action tendencies, particularly acting out and aggressive behavior, in adults and children
- 10 cards (1 blank)

Apperceptive Personality Test (APT)
- Holmstrom et. Al.
- attempt to address the criticisms of TAT
- introduced objectivity in scoring system
- 8 cards include male and female of different ages and minority group members
- testtakers will respond to a series of multiple choice questions after storytelling

Word Association Test (WAT)
- Rapaport et. Al.
- presentation of a list of stimulus words, assessee responds verbally or in writing the first thing that comes into their minds

Rotter Incomplete Sentences Blank (RISB)
- Julian Rotter & Janet Rafferty
- Grade 9 to Adulthood
- most popular SCT

SACK's Sentence Completion Test (SSCT)
- Joseph Sacks and Sidney Levy
- 12 years old and older
- asks respondents to complete 60 questions with the first thing that comes to mind across four areas: Family, Sex, Interpersonal, Relationships and Self concept

Bender-Gestalt Visual Motor Test
[B]
- Lauretta Bender
- 4 years and older
- consists of a series of durable template cards, each displaying a unique figure, then they are asked to draw each figure as he or she observes it
- provides interpretative information about an individual's development and neuropsychological functioning
- reveals the maturation level of visuomotor perceptions, which is associated with language ability and various functions of intelligence

House-Tree-Person Test (HTP)
- John Buck and Emmanuel Hammer
- 3 years and up
- measures aspects of a person's personality through interpretation of drawings and responses to questions
- can also be used to assess brain damage and general mental functioning
- measures the person's psychological and emotional functioning
- The house reflects the person's experience of their immediate social world
- The tree is a more direct expression of the person's emotional and psychological sense of self
- The person is a more direct reflection of the person's sense of self

Draw-A-Person Test (DAP)
- Florence Goodenough
- 4 to 10 years old
- a projective drawing task that is often utilized in psychological assessments of children
- Aspects such as the size of the head, placement of the arms, and even things such as if teeth were drawn or not are thought to reveal a range of personality traits
- Helps people who have anxieties taking tests (no strict format)
- Can assess people with communication problems
- Relatively culture free
- Allow for self-administration

Kinetic Family Drawing
- Burns & Kaufman
- derived from Hulses' FDT "doing something"

Controlled Oral Word Association Test (COWAT)
- measures spontaneous production of words belonging to the same category or beginning with some designated letter

Clinical & Counseling Tests

Millon Clinical Multiaxial Scale-IV (MCMI-IV)
Theodore Millon
- 18 years old and above
- standardized, self-report questionnaire that assesses a wide range of information related to a client's personality, emotional adjustment, and attitude toward taking tests
- for diagnosing and treatment of personality disorders
- exaggeration of polarities results to maladaptive behavior

Source: Cohen & Swerdlik (2018), Kaplan & Saccuzzo (2018), Groth & Wright (2016), Psych Pearls

Validity/Modifying Indices:
1. Invalidity (V) – person answers randomly
2. Inconsistency (W) – expected to be answered both in same direction
3. Disclosure (X) – designed to measure whether a client's responses were open and revealing as opposed to defensive and secretive
4. Desirability (Y) – measure of defensive responding
5. Debasement (Z) – extent to which a person is describing himself in negative, pathological terms

Clinical Personality Patterns:
1. Schizoid
2. Avoidant
3. Melancholic
4. Dependent
5. Histrionic
6. Turbulent
7. Narcissistic
8. Antisocial
9. Sadistic
10. Compulsive
11. Negativistic
12. Masochistic

Severe Personality Pathology:
1. Schizotypal
2. Borderline
3. Paranoid

Clinical Syndromes:
1. Generalized Anxiety
2. Somatic Symptom
3. Bipolar Spectrum
4. Persistent Depression
5. Alcohol Use
6. Drug Use
7. Posttraumatic Stress

Severe Clinical Syndromes:
1. Schizophrenic Spectrum
2. Major Depression
3. Delusional

Validity Indicators:
1. *Random Responding:* scores of one or more on the three items of the MCMI-IV Invalidity scale (items 49, 98, 160)
2. *Underreporting of Difficulties (Faking Good):* low scores on Disclosure (X) and Debasement (Z), high score on Desirability (Y)
3. *Fake Bad:* high score on Disclosure (X) and high score in Debasement (Z)

Beck Depression Inventory (BDI-II)
- Aaron Beck
- 13 to 80 years old
- 21-item self-report that taps Major Depressive symptoms accdg. to the criteria in the DSM
- *Beck's Anxiety Inventory (BAI)*

MacAndrew Alcoholism Scale (MAC & MAC-R)
- from MMPI-II
- Personality & attitude variables thought to underlie alcoholism

California Psychological Inventory (CPI-III)
- attempts to evaluate personality in normally adjusted individuals
- has validity scales that determines faking bad and faking good
- interpersonal style and orientation, normative orientation and values, cognitive and intellectual function, and role and personal style
- has special purpose scales, such as managerial potential, work orientation, creative temperament, leadership potential, amicability, law enforcement orientation, tough-mindedness

Rosenberg Self-Esteem Scale
- measures global feelings of self-worth
- 10-item, 4 point likert scale
- used with adolescents

Dispositional Resilience Scale (DRS)
- measures psychological hardiness defined as the ability to view stressful situations as meaningful, changeable, and challenging

Ego Resiliency Scale-Revised
- measure ego resiliency or emotional intelligence

HOPE Scale
- developed by Snyder
- *Agency:* cognitive model with goal driven energy
- *Pathway:* capacity to contrast systems to meet goals
- good measure of hope for traumatized people
- positively correlated with health psychological adjustment, high achievement, good problem solving skills, and positive health-related outcomes

Satisfaction with Life Scale (SWLS)
- overall assessment of life satisfaction as a cognitive judgmental process

Positive and Negative Affect Schedule (PANAS)

- measure the level of positive and negative emotions a test taker has during the test administration

Pictorial Self-Concept Scale for Children (PSC)
- Jack Joseph
- 3 yrs to 13 yrs old
- Allows clinician to measure self-concept in children as young as 3
- Identifies children with negative self-appraisal put them at risk for academic and behavioral difficulties
- Let youngster respond using pictures rather than words
- Children are shown pairs of illustrations representing common self-appraisal situations and are asked to choose between a picture representing negative and positive self-concept
- Can be used to evaluate psychological and educational interventions

Halstead-Reitan Neuropsychological Battery (HRNB)
- Ward Halstead & Ralph Reitan
- 15 yrs old and older
- set of tests used to diagnose localize brain damage by providing a comprehensive assessment of cognitive functioning

Five Core Subtests:
1. Category Test
2. Tactual Performance Test
3. Seashore Rhythm Test
4. Speech Sounds Perception
5. Finger Tapping Test

Five Optional Subtests:
1. Trail Making Test
2. Reitan Indiana Aphasia Screening Test
3. Reitan-Klove Sensory Perceptual Examination
4. Grip Strength Test
5. Lateral Dominance Examination

Neuropsychological Impairment Scale (NIS) [C]
- William E. O'Donell et. Al.
- 18 years old – 88 yrs old
- self-report, observer report, and senior interview
- identify neuropsychological symptoms and defiency
- produces one Global Measure of Impairment (GMI)

Supports Intensity Scale (SIS-A/SIS-C)
- evaluates the support needed by an individual
- focuses on the types and intensities of supports needed to enable an individual to lead a normal independent life

6 Categories:
1. Home Living
2. Community Living
3. Life-Long Learning
4. Employment
5. Health and Safety
6. Social Activities

Self-Directed Search (SDS)
- career assessment and exploration tool that matches your aspirations, activities, and talents to the career choices and educational opportunities that fit you best

Eating Disorder Inventory (EDI-3)
- self-report measures of constructs shown to be clinically relevant in individuals with eating disorders

12 Primary Scales
Drive for Thinness
Bulimia
Body Dissatisfaction
Low Self-Esteem
Personal Alienation
Interpersonal Insecurity
Interpersonal Alienation
Interoceptive Deficits
Emotional Dysregulation
Perfectionism
Asceticism
Maturity Fears

Vineland Adaptive Behavior Scales (VABS-II)
- leading instrument for supporting the diagnosis of intellectual and developmental disabilities

Domains:
1. Communication
2. Daily Living Skills
3. Socialization

Supplemental Scales:
1. Motor Skills
2. Maladaptive Behaviors

Kiddie Schedule of Affective Disorders and Schizophrenia-Present and Lifetime Version (K-SADS-PL)
- assesses hypomania, cyclothymia, and BD
- also combines dimensional and categorical assessment approaches to diagnose current and past episodes of psychopathology in children and adolescents

Child Behavior Checklist (CBCL)

Source: Cohen & Swerdlik (2018), Kaplan & Saccuzzo (2018), Groth & Wright (2016), Psych Pearls

- 6 yrs old to 18 yrs old
- Parent or other close relatives of the child complete the form regarding the behavior of the client
- 118 items regarding behavior problems
- Teacher's Report Form and Youth Self-Report Form
- can assist in identifying BD in children

Subscales:
Aggressive Behavior
Anxious/Depressed
Attention Problems
Rule-Breaking Behavior
Social-Problems
Somatic Complaints
Thought Problems
Withdrawn/Depressed

Revised Child Anxiety and Depression Scale (RCADS)
- 8 yrs old to 18 yrs old
- 47-item, youth self-report questionnaire
- *Revised Child Anxiety and Depression Scale – Parent Version (RCADS-P)*

Subscales:
Anxiety Disorder
Social Phobia
GAD
Panic Disorder
OCD
Low Mood

Screen for Child Anxiety Related Emotional Disorders (SCARED)
- measure widely used to assess childhood anxiety based on parent and child support
- Child and parent

Spence Children's Anxiety Scale (SCAS)
- 8 yrs old to 15 yrs old
- assess severity of anxiety symptoms in children
- Child and Parent

Subscales:
Separation Anxiety
Social Phobia
OCD
Panic/Agoraphobia
GAD
Fears of Physical Injury

Preschool Anxiety Scale (PAS)
- 2 ½ yrs old to 6 ½ yrs old
- completed by parent/guardian
- assesses the anxiety in children
- intended to provide an indicator of the number and severity of anxiety symptoms experienced by younger children

Subscales
GAD
Social Anxiety
OCD
Physical Injury Fears
Separation Anxiety

Children's Yale-Brown Obsessive-Compulsive Scale (CY-BOCS)
- 6 yrs old to 17 yrs old
- designed to rate the severity of OC symptoms in children and adolescents
- clinician-rated

Dimensional Yale-Brown Obsessive-Compulsive Scale (DY-BOCS)
- assesses the presence and severity of obsessive compulsive-symptoms

Children's Obsessional Compulsive Inventory – Revised Self Report (ChOCI-R-S)
- 32-item self-report measure assessing the presence and severity of obsessive compulsive disorder (OCD) in children and adolescents aged 7-17 years

Vineland Adaptive Behavior Scale (VABS)
[B]
- most popular measures of adaptive behavior in children
- 0-90 years old

Adaptive Behavior Assessment System (ABAS-III)
- used to assess the functional skills necessary for the daily living
- 0-89 years old

7-Minute Screen
- identify patients with Alzheimer's Disease

Strengths and weaknesses of assessment tools (2)

Pros	Cons
Test	
- can gather a sample of behavior objectively with lesser bias - flexible, can be verbal or nonverbal	- In crisis situations when relatively rapid decisions need to be made, it can be impractical to take the time required to administer and interpret tests
Interview	
- can take note of verbal and nonverbal cues	- sometimes, due to negligence of interviewer

Source: Cohen & Swerdlik (2018), Kaplan & Saccuzzo (2018), Groth & Wright (2016), Psych Pearls

- flexible - time and cost effective - both structured and unstructured allows clinicians to place a wider, more meaningful context - can also be used to help predict future behaviors interviews allow - clinicians to establish rapport and encourage client self-exploration.	and interviewee, it can miss out important information - interviewer's effect on the interviewee - various error such as halo effect, primacy effect, etc. - interrater reliability - interviewer bias
Portfolio	
- provides comprehensive illustration of the client which highlights the strengths and weaknesses	- can be very demanding - time consuming
Observation	
- flexible - suitable for subjs that cannot be studied in lab setting - more realistic - affordable - can detect patterns	- For private practitioners, it is typically not practical or economically feasible to spend hours out of the consulting room observing clients as they go about their daily lives - lack of scientific control, ethical considerations, and potential for bias from observers and subjects - unable to draw cause-and-effect conclusions - lack of control - lack of validity - observer bias
Case History	
- can fully show the experience of the observer in the program - shed light on an individual's past and current adjustment as well as on the events and circumstances that may have contributed to any changes in adjustment	- cannot be used to generalize a phenomenon
Role Play	
- encourages individuals to come together to find solutions and to get to	- may not be as useful as the real thing in all situations - time-consuming
know how their colleagues think - group can discuss ways to potentially resolve the situation and participants leave with as much information as possible, resulting in more efficient handling of similar real-life scenarios	- expensive - inconvenient to assess in a real situation - While some employees will be comfortable role playing, they're less adept at getting into the required mood needed to actually replicate a situation

Test Administration, Scoring, Interpretation and Usage (20)

Detect Errors and impacts in Test

Issues in Intelligence Testing

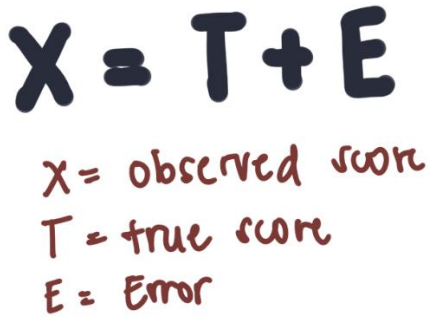
1. **Flynn Effect –** progressive rise in intelligence score that is expected to occur on a normed intelligence test from the date when the test was first normed
   - Gradual increase in the general intelligence among newborns
   - *Frog Pond Effect:* theory that individuals evaluate themselves as worse when in a group of high-performing individuals
2. **Culture Bias of Testing**
   - *Bias In Testing:* presence of systematic error in the measurement of certain factors
   - *Culture-Free:* attempt to eliminate culture so nature can be isolated
   - Impossible to develop bec culture is evident in its influence since birth or an individual and the interaction between nature and nurture is cumulative and not relative
   - *Culture Fair:* minimize the influence of culture with regard to various aspects of the evaluation procedures
   - Fair to all, fair to some cultures, fair only to one culture
   - *Culture Loading:* the extent to which a test incorporates the vocabulary concepts traditions, knowledge etc. with particular culture

Test Equivalence
Linguistic: wording and content - whether the test has been translated accurately - *Back-Translation:* once translated to a certain language, it is translated back to the original language
Conceptual: construct has the same meaning - requires constructs to have the same meaning in various cultures
Metric: same psychometric feature across different groups/culture

Errors: Reliability

Source: Cohen & Swerdlik (2018), Kaplan & Saccuzzo (2018), Groth & Wright (2016), Psych Pearls

- **Classical Test Theory (True Score Theory)** – score on ability tests is presumed to reflect not only the testtaker's true score on the ability being measured but also the error
  - *Error:* refers to the component of the observed test score that does not have to do with the testtaker's ability
  - Errors of measurement are random



  - *The greater number of items, the higher the reliability*
  - Factors that contribute to inconsistency: *characteristics* of the individual, test, or situation, which have nothing to do with the attribute being measured, but still affect the scores
- **Error Variance** – variance from irrelevant random sources

Measurement Error – all of the factors associated with the process of measuring some variable, other than the variable being measured
- difference between the observed score and the true score
- *Positive:* can increase one's score
- *Negative:* decrease one's score
- Sources of Error Variance:

a. **Item Sampling/Content Sampling**
b. **Test Administration**
c. **Test Scoring and Interpretation**

Random Error – source of error in measuring a targeted variable caused by unpredictable fluctuations and inconsistencies of other variables in measurement process (e.g., noise, temperature, weather)

Systematic Error – source of error in a measuring a variable that is typically constant or proportionate to what is presumed to be the true values of the variable being measured
- has consistent effect on the true score
- SD does not change, the mean does

  - Error variance may increase or decrease a test score by varying amounts, consistency of test score, and thus, the reliability can be affected

Test-Retest Reliability

Error: *Time Sampling*
- the longer the time passes, the greater likelihood that the reliability coefficient would be insignificant
- *Carryover Effects:* happened when the test-retest interval is short, wherein the second test is influenced by the first test because they remember or practiced the previous test = inflated correlation/overestimation of reliability
- *Practice Effect:* scores on the second session are higher due to their experience of the first session of testing
- test-retest with longer interval might be affected of other extreme factors, thus, resulting to low correlation
- target time for next administration: at least two weeks

Parallel Forms/Alternate Forms Reliability

Error: *Item Sampling* (Immediate), *Item Sampling changes over time* (delayed)
- *Counterbalancing:* technique to avoid carryover effects for parallel forms, by using different sequence for groups
- most rigorous and burdensome, since test developers create two forms of the test
- main problem: difference between the two tests
- test scores may be affected by motivation, fatigue, or intervening events
- create a large set of questions that address the same construct then randomly divide the questions into two sets

Internal Consistency (Inter-Item Reliability)

Error: *Item Sampling Homogeneity*

Split-Half Reliability

Error: *Item sample: Nature of Split*

Inter-Scorer Reliability

Error: *Scorer Differences*

- **Standard Error of Measurement** – provide a measure of the precision of an observed test score
  - Standard deviation of errors as the basic measure of error
  - Index of the amount of inconsistent or the amount of the *expected* error in an individual's score
  - Allows to quantify the extent to which a test provide accurate scores
  - Provides an estimate of the amount of error inherent in an observed score or measurement
  - Higher reliability, lower SEM
  - Used to estimate or infer the extent to which an observed score deviates from a true score
  - **Standard Error of a Score**
  - *Confidence Interval:* a range or band of test scores that is likely to contain true scores

Source: Cohen & Swerdlik (2018), Kaplan & Saccuzzo (2018), Groth & Wright (2016), Psych Pearls

- **Standard Error of the Difference** – can aid a test user in determining how large a difference should be before it is considered statistically significant
- **Standard Error of Estimate** – refers to the standard error of the difference between the predicted and observed values
- **Four Possible Hit and Miss Outcomes**
  1. **True Positives (Sensitivity)** – predict success that does occur
  2. **True Negatives (Specificity)** – predict failure that does occur
  3. **False Positive (Type 1)** – success does not occur
  4. **False Negative (Type 2)** – predicted failure but succeed

Errors due to Behavioral Assessment

1. **Reactivity** – when evaluated, the behavior increases
- Hawthorne Effect

2. **Drift** – moving away from what one has learned going to idiosyncratic definitions of behavior
- subjects should be retrained in a point of time
- *Contrast Effect:* cognitive bias that distorts our perception of something when we compare it to something else, by enhancing the differences between them

3. **Expectancies** – tendency for results to be influenced by what test administrators expect to find
- *Rosenthal/Pygmalion Effect:* Test administrator's expected results influences the result of the test
- *Golem Effect:* negative expectations decreases one's performance

4. **Rating Errors** – intentional or unintentional misuse of the scale
- *Leniency Error:* rater is lenient in scoring (Generosity Error)
- *Severity Error:* rater is strict in scoring
- *Central Tendency Error:* rater's rating would tend to cluster in the middle of the rating scale
- *Halo Effect:* tendency to give high score due to failure to discriminate among conceptually distinct and potentially independent aspects of a ratee's behavior
- snap judgement on the basis of positive trait
- *Horn Effect:* Opposite of Halo Effect
- One way to overcome rating errors is to use rankings

5. **Fundamental Attribution Error** – tendency to explain someone's behavior based on internal factors such as personality or disposition, and to underestimate the influence the external factors have on another person's behavior, blaming it on the situation
- *Barnum Effect:* people tend to accept vague personality descriptions as accurate descriptions of themselves (*Aunt Fanny Effect*)

- **Bias** – factor inherent in a test that systematically prevents accurate, impartial measurement
  - Prejudice, preferential treatment
  - Prevention during test dev through a procedure called **Estimated True Score Transformation**
- **Social Desirability** – tendency to say good things about yourself or to mark items that you believe will be approved by the examiner

Ethical Principles and Standards of Practice (19)

- Assessment should only be conducted in the context of a clearly defined professional relationship, thus, nature, purpose, and conditions of the relationship must be discussed and agreed on
- It is the examiner's responsibility to recognize the possible influences he or she may exert on the client and to optimize the level of rapport
- It is particularly important to assess the degree of the client's motivation and his or her overall level of anxiety
  - There could be times that testing must be halted due to the client's emotional state
- The more the examiner likes the client, the more likely he or she will be to score an ambiguous response in a direction favorable to the client
- If mistakes were made, they should do something to correct or minimize the mistakes
- If an ethical violation made by another psychologist was witnessed, they should resolve the issue with informal resolution, as long as it does not violate any confidentiality rights that may be involved
- If informal resolution is not enough or appropriate, referral to state or national committees on professional ethics, state licensing boards, or the appropriate institutional authorities can be done. Still, confidentiality rights of the professional in question must be kept.

- Failure to cooperate in ethics investigation itself, is an ethics violation, unless they request for deferment of adjudication of an ethics complaint
- Psychologists must file complaints responsibly by checking facts about the allegations
- Psychologists DO NOT deny persons employment, advancement, admissions, tenure or promotion based solely upon their having made or their being the subject of an ethics complaint
  - Just because they are questioned by the ethics committee or involved in an on-going ethics investigation, they would be discriminated or denied advancement
  - Unless the outcome of the proceedings are already considered
- Psychologists should do their services within the boundaries of their competence, which is based on the amount of training, education, experience, or consultation they had
- When they are tasked to provide services to clients who are deprived with mental health services (e.g., communities far from the urban cities), however, they were still not able to obtain the needed competence for the job, they could still provide services AS LONG AS they make reasonable effort to obtain the competence required, just to ensure that the services were not denied to those communities
- During emergencies, psychologists provide services to individuals, even though they are yet to complete the competency/training needed just to ensure that services were not denied. However, the services are discontinued once the appropriate services are available
- Any consent involves a clear explanation of what procedures will occur, the relevance of the testing, and how the results will be used

Privacy	**Confidentiality**
- actions of honoring an individual's right to control who has access for him/herself and how their information is shared publicly - personal - right	- refers to actions of keeping documents, files, and other data away from people or entities that are not meant to see/hear them - relational - agreement

- Psychologists should discuss the limits of confidentiality, uses of the information that would be generated from the services to the persons and organizations with whom they establish a scientific or professional relationships
- Before recording voices or images, they must obtain permission first from all persons involved or their legal rep
- Only discuss confidential information with persons clearly concerned/involved with the matters
- Disclosure is allowed with appropriate consent
  - No consent is not allowed UNLESS mandated by the law
- No disclosure of confidential information that could lead to the identification of a client unless they have obtained prior consent or the disclosure cannot be avoided
  - Only disclose necessary information
- Exemptions to disclosure:
  - ✓ If the client is disguised/identity is protected
  - ✓ Has consent
  - ✓ Legally mandated
- Psychologists can create public statements as long as they would be responsible for it
  - They cannot compensate employees of the media in return for publicity in a news item
  - Paid Advertisement must be clearly recognizable
  - when they are commenting publicly via internet, media, etc., they must ensure that their statement are based on their professional knowledge in accord with appropriate psych literature and practice, consistent with ethics, and do not indicate that a professional relationship has been established with the recipient
- Must provide accurate information and obtain approval prior to conducting the research
- Informed consent is required, which include:
  - ✓ Purpose of the research
  - ✓ Duration and procedures
  - ✓ Right to decline and withdraw
  - ✓ Consequences of declining or withdrawing
  - ✓ Potential risks, discomfort, or adverse effects
  - ✓ Benefits
  - ✓ Limits of confidentiality
  - ✓ Incentives for participation
  - ✓ Researcher's contact information
- Permission for recording images or vices are needed unless the research consists of solely naturalistic observations in public places, or research designed includes deception
  - Consent must be obtained during debriefing
- Dispense or Omitting Informed consent only when:
  1. Research would not create distress or harm
     - Study of normal educational practices conducted in an educational settings

- Anonymous questionnaires, naturalistic observation, archival research
- Confidentiality is protected

2. Permitted by law

- Avoid offering excessive incentives for research participation that could coerce participation
- DO not conduct study that involves deception unless they have justified the use of deceptive techniques in the study
  - Must be discussed as early as possible and not during the conclusion of data collection
- They must give opportunity to the participants about the nature, results, and conclusions of the research and make sure that there are no misconceptions about the research
- Must ensure the safety and minimize the discomfort, infection, illness, and pain of animal subjects
  - If so, procedures must be justified and be as minimal as possible
  - During termination, they must do it rapidly and minimize the pain
- Must no present portions of another's work or data as their own
  - Must take responsibility and credit, including authorship credit, only for work they have actually performed or to which they have substantially contributed
  - Faculty advisors discuss publication credit with students as early as possible
- After publishing, they should not withhold data from other competent professionals who intends to reanalyze the data
  - Shared data must be used only for the declared purpose
- **RA 9258 –** Guidance and Counseling Act of 2004
- **RA 9262 –** Violence Against Women and Children
- **RA 7610 –** Child Abuse
- **RA 9165 –** Comprehensive Dangerous Drugs Act of 2002
- **RA 11469 –** Bayanihan to Heal as One Act
- **RA 7277 –** Magna Carta for Disabled Persons
- **RA 11210 –** Expanded Maternity Leave Law
- **RA 11650 –** Inclusive Education Law
- **RA 10173 –** Data Privacy Act
- **House Bill 4982 –** SOGIE Bill
- **Art. 12 of Revised Penal Code –** Insanity Plea

Additional Information

- **Thurstone: Seven Primary Mental Abilities**
  - Verbal Comprehension
  - Word Fluency
  - Number Facility
  - Spatial Visualization
  - Associative Memory
  - Perceptual Speed
  - Reasoning
- **Cattell: Crystallized and Fluid Intelligence**
- **Spearman: Two-Factor Theory of Intelligence**
  - General (g) Intelligence
  - Specific (s) Ability
- **Gardner: Theory of Multiple Intelligences**
  - Visual-Spatial
  - Linguistic-Verbal
  - Logical-Mathematical
  - Body-Kinesthetic
  - Musical
  - Interpersonal
  - Intrapersonal
  - Naturalistic

Malingering – the deliberate feigning of an illness or disability to achieve a particular desired outcome
Faking Bad (Overreporting) - appear worse than actually is the case
Faking Good (Underreporting) - attempt to appear better than is actually the case - Intentional (Social Desirability) or Unintentional (Impression Management)
Nay-Saying – answering question negatively of their content, which can distort results of surveys, questionnaires, and similar instruments - *Non-acquiescence*
Yea-Saying – answering question positively regardless of their content, which can distort the results of surveys - *Acquiescence*
Self-Deception – process of denying or rationalizing away the relevance, significance, or importance of opposing evidence and logical argument

- **Jungian Typology:**

Extraverted Thinking – principle, idealistic, objective, rational
Introverted Thinking – influenced by ideas, independent, often fearful of intimacy
Extraverted Feeling – adaptive, relating well to the external environment
Introverted Feeling – sympathetic, pleases others, may be dependent, reserved
Extraverted Sensation – realistic, concrete, pleasant, and friendly
Introverted Sensation – calm, passive, restrained, controlled, and controlling
Extraverted Intuition – enterprising, outgoing, can be irresponsible

Introverted Intuition – mystical, dreamer, and artist, can be obsessive

- **Types of Interpretation:**

Concrete – limited to the subtest and subscale scores and does not draw conclusions beyond the scores
Mechanical – pattern of subscales; significant difference between scores
Individualized – interpreting the results in the context of the larger picture

- **16 PF Factors**

Global Factors	Cont. Primary Factors
Extraversion (4)	Warmth Liveliness Social Boldness Privateness Self-Reliance
Independence (4)	Dominance Social Boldness Vigilance Openness to Change
Tough-Mindedness (6)	Warmth Sensitivity Abstractedness Openness to Change
Self-Control (4)	Liveliness Rule-Consciousness Abstractedness Perfectionism
Anxiety (5)	Emotional Stability Vigilance Apprehension Tension

- **Interpretation of Assessment Information:**

Intuitive Approach – no identifiable basis, based on feelings and instincts, and have no empirical explanation
Authoritative Approach – based on experts or well-known psychologists
Empirical/Conceptual – derived from researches

- **Function of Tests:**

Additive Function – measuring different domains to broaden understanding of the person being examined
Confirmatory Function – two or more measures of the same domain are placed in a test battery to strengthen initial impressions
Complementary Function – multiple measures of the same dome yield different rather than confirmatory results

- **Mental Status Exam** – appropriate before assessment to determine the appropriateness of more formal psychological testing

Appearance – client's clothing, posture, gestures, speech, personal care/hygiene, and any unusual physical features
Relatedness – facial expressions, eye contact, activity level, degree of cooperation, notable physical characteristics, and attentiveness
Speech/Language – often proxies for their thought processes, as they relate to the primary mode of communicating thoughts to the outside world - help clinicians determine the possibility of poor or exceptional cognitive functioning, focus, and confusion and possible thought disorder - evaluating how well individuals understand language, as evidenced by responding appropriately to directions and conversations (receptive language)
Affect/Mood (Feeling) – client might be cold or warm, distant or close, labile, or have flat affect - *Mood:* dominant emotion reported during the interview - *Affect:* client's outwardly projected range of emotions
Thought Process
Thought Content – can be coherent, spontaneous, and comprehensible or may contain unusual features
Memory
Attention/Concentration
Alertness/Orientation – they know who they are, where they are, and when current and past events have occurred or are occurring - *Sensorium:* refers to how intact their physical sensory processes are to receiving and integrating information (hearing, smell, vision, and touch and might range from being clouded to clear)
Judgment/Planning – vary how they provide for themselves, evaluate risks, and make plans
Insight – why they believe they were referred for evaluation and, in a wider context, they attitudes towards their cultures

- **Types of Decisions made through Psychological Tests:**

Individual – testtakers use their scores to make decisions about themselves
Institutional – made by another entity about a person's score
Comparative – made by comparing the test scores of a number of people to see who has the best score
Absolute – decisions made by other by looking who has the minimum score needed to qualify

Source: Cohen & Swerdlik (2018), Kaplan & Saccuzzo (2018), Groth & Wright (2016), Psych Pearls

end

Congratulations *for reaching the end of this reviewer! <3*

Remember to ***take rest*** *if you need to and* ***be less harsh*** *to yourself.* ***Reward yourself****, you deserve it. You* ***can never learn everything*** *but at least* ***you still did learn something****.* ***Progress is progress.*** *The most important thing is* ***you will get there****!*

Claim that license*!*

Congratulations*, Future* ***RPm****!*
