PSYCHOLOGICAL ASSESSMENT

PmLE | **Main References: Cohen & Swerdlik, Kaplan & Sacuzzo, Miller & Lovler, Urbina**

TESTING VS. ASSESSMENT		
Objective	Quantify a construct	Answer a referral question
Focus	Nomothetic approach	Idiographic approach
Process	Individual or group	Individual only
Outcome	Test score or psychometric report	Psychological report
Source	Test takers only	Collateral sources
Evaluator	Tester is not the key	Assessor is the key
Duration	Shorter	Longer
Cost	Inexpensive	Expensive
Qualification	RPm	RPsy

PSYCHOLOGICAL ASSESSMENT – gathering and integration of psychology-related data for the purpose of making a psychological evaluation that is accomplished through the use of tools, tests, interviews, case studies, behavioral observation, and specially designed apparatuses

- **Dynamic Psychological Assessment –** interactive approach; evaluation-intervention-evaluation ("sandwich method")
- **Collaborative Psychological Assessment –** assessor and assessee may work as "partners" from the initial contact to the final feedback
- **Therapeutic Psychological Assessment –** therapeutic self-discovery and new understandings throughout the process

Varieties of Assessment

- ➔ **Educational Assessment –** use of tests and other tools to evaluate abilities and skills relevant to success or failure in a pre-/school context (e.g., intelligence tests, achievement tests, reading comprehension tests)
- ➔ **Retrospective Assessment –** use of evaluative tools to conclude psychological aspects of a person as they existed at some point in time before the assessment
- ➔ **Remote Assessment –** use of psychological tools to gather data and draw conclusions about subjects who are not in physical proximity to the person or people conducting the evaluation
- ➔ **Ecological Momentary Assessment –** "in the moment" evaluation of specific problems and related cognitive and behavioral variables at the very time and place that they occur

PSYCHOLOGICAL TESTING – process of measuring psychological-related variables by means of devices or procedures designed to obtain a sample of behavior

- **Psychological Test –** a device or procedure designed to measure variables related to psychology

Variables of Testing

- ➔ **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 Procedure –** individual basis or group administration
- ➔ **Score –** code or summary of statement, usually but not necessarily numerical in nature, that reflects an evaluation of performance on a test
- ➔ **Scoring –** the process of assigning scores to performances
- ➔ **Cut Score –** a reference point, usually numerical, derived by judgment and used to divide a set of data into two or more classifications
- ➔ **Psychometric Soundness –** technical quality
- ➔ **Psychometric –** the science of psychological measurements
- ➔ **Psychometrician –** a professional who uses, analyzes, and interprets psychological test data

PSYCHOMETRIC PROPERTIES AND PRINCIPLES (39)

PSYCHOMETRIC PROPERTIES – refer to the technical qualities and characteristics of psychological tests and assessments that determine their scientific soundness and practical usefulness

- These properties are crucial for ensuring that a test is measuring what it intends to measure, that its results are consistent, and that the scores can be meaningfully interpreted

TOS-BASED REVIEWERS by Anne (rpm_anne, eireandtheclouds@gmail.com) IS ENTIRELY FOR FREE | NOT FOR SALE PLEASE!

PSYCHOLOGICAL ASSESSMENT

PmLE | Main References: Cohen & Swerdlik, Kaplan & Sacuzzo, Miller & Lovler, Urbina

RELIABILITY

RELIABILITY – consistency in measurement

- **Reliability Coefficient –** an index of reliability, a proportion that indicates the ratio between the true score variance on a test and the total score, ranging from 0 to 1
- Synonymous with consistency or dependability
- The greater the proportion of the total variance attributed to the true variance, the more reliable the test
- True Score formula:

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

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

Goals of Reliability

➔ Estimate errors in psychological measurement
➔ Device techniques to improve testing so errors are reduced

$$\frac{TRUE\ VARIANCE}{TRUE\ SCORE\ VARIANCE + ERROR\ VARIANCE}$$

Classical Test Theory – a score on an ability is presumed to reflect not only the test taker's true score on the ability being measured but also the *error*

- **Error –** the component of observed test score that does not have to do with the test taker's ability
  - X = T (true score) + E (error)
- **Variance –** useful in describing sources of test score variability
  - **True Variance –** variance from true difference
  - **Error Variance –** variance from irrelevant random sources

Measurement Error – all of the factors associated with the process of measuring some variables, other than variables being measured

- **Random Error –** source of error in measuring a targeted variable caused by unpredictable fluctuations and inconsistencies of other variables in the measurement process. *Unavoidable*
- **Systematic Error –** source of error in measuring a variable that is typically constant or proportionate to what is presumed to be the true value of the variable being measured. *Avoidable if corrected*

Sources of Error

- **Test Construction**
  - **Item Sampling (Content Sampling) –** refers to variation among items within a test as well as to variation among items between tests.
  - The extent to which a testtaker's score is affected by the content sampled on a test and by the way the content is sampled (that is, the way in which the item is constructed) is a source of error variance.
- **Test Administration**
  - **Test Environment** (room temperature, level of lighting, and the amount of ventilation and noise)
  - **Test taker variable** (emotional problems, physical discomfort, lack of sleep, effects of drug, formal learning experiences, casual life experiences, therapy, illness, and changes in mood or mental state)
  - **Examiner-related variables** (examiner's physical appearance and demeanor, nonverbal gestures, and professionalism)
- **Test Scoring and Interpretation**
  - Scorers and scoring systems are potential sources of error variance. A test may employ objective-type items amenable to computer scoring of well-documented reliability.
  - If subjectivity is involved in scoring, then the scorer (or rater) can be a source of error variance.
  - The element of subjectivity in scoring may be much greater in the administration of certain nonobjective-type personality tests, tests of creativity (such as the block test just described), and certain academic tests (such as essay examinations)

2 Types of Reliability

➔ **Internal Reliability –** has consistency within itself (e.g., "I hate Mondays" and "I don't like Mondays" are worded differently but have one meaning)
➔ **External Consistency –** how results compare to results between individuals and across time (e.g., pretest and posttest, interrater difference)

A | Reliability Estimates

Test-Retest Reliability – is an estimate of reliability obtained by correlating pairs of scores from the same people on two different administrations of the same test.

- The test-retest measure is appropriate when evaluating the reliability of a test that purports to measure something that is relatively stable over time, such as a personality trait.

eireandtheclouds
STUDY NOTES BY ANNE
THIS IS NOT FOR SALE.
ANY FORMS OF REPRODUCTION
IS STRICTLY PROHIBITED.
eireandtheclouds@gmail.com | x: rpm_

- The longer the time that passes, the greater the likelihood that the reliability coefficient will be lower.
- **Coefficient of Stability –** the estimate of test-retest reliability when the interval between testing is greater than 6 months.
- An estimate of test-retest reliability may be most appropriate in gauging the reliability of tests that employ outcome measures such as reaction time or perceptual judgments (including discriminations of brightness, loudness, or taste).
- Even in measuring variables such as these, and even when the time period between the two administrations of the test is relatively small, various factors (such as experience, practice, memory, fatigue, and motivation) may intervene and confound an obtained measure of reliability.
- **Carryover Effects –** happened when test-retest interval is short, wherein the second test is influenced by the first test because they remember or practiced the previous test
- **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 highlighted as confused
- **Test Wiseness –** might inflate the abilities of test takers
- **Mortality –** problems in absence in second session (just remove the first test of the absents)

Parallel- and Alternate-Forms Reliability

- **Parallel Forms Reliability –** refers to an estimate of the extent to which item sampling and other errors have affected test scores on versions of the same test when, for each form of the test, the means and variances of observed test scores are equal.
    - **Parallel Forms –** of a test exist when, for each form of the test, the means and the variances of observed test scores are equal.
    - In theory, the means of scores obtained on parallel forms correlate equally with the true score.
    - More practically, scores obtained on parallel tests correlate equally with other measures.
- **Alternate Forms Reliability –** refers to an estimate of the extent to which these different forms of the same test have been affected by item sampling error, or other error.
    - **Alternate Forms –** are simply different versions of a test that have been constructed so as to be parallel.
        - Although they do not meet the requirements for the legitimate designation "parallel," alternate forms of a test are typically designed to be equivalent with respect to variables such as content and level of difficulty.
    - Obtaining Alternate-Forms Reliability and Parallel-Forms Reliability (similar in test-retest)
        - Two test administrations with the same group are required.
        - Test scores may be affected by factors such as motivation, fatigue, or intervening events such as practice, learning, or therapy (although not as much as when the same test is administered twice).
- **Counterbalancing –** a technique to avoid carryover effects for parallel forms, by using different sequences for groups.

Internal Consistency (Inter-Item Consistency) – refers to the degree of correlation among all the items on a scale.

- **Homogeneity –** a measure of inter-item consistency is calculated from a single administration of a single form of a test.
    - Tests are said to be homogeneous if they contain items that measure a single trait.
- **Heterogeneity –** describes the degree to which a test measures different factors.
    - A heterogeneous (or nonhomogeneous) test is composed of items that measure more than one trait.
- The more homogeneous a test is, the more inter-item consistency it can be expected to have.
- Because a homogeneous test samples a relatively narrow content area, it is to be expected to contain more inter-item consistency than a heterogeneous test.
- Test homogeneity is desirable because it allows relatively straightforward test-score interpretation.
- Test takers with the same score on a homogeneous test probably have similar abilities in the area tested.
- Test takers with the same score on a more heterogeneous test may have quite different abilities.
- Although a homogeneous test is desirable because it so readily lends itself to clear interpretation, it is often an insufficient tool for measuring multifaceted psychological variables such as intelligence or personality.
- One way to circumvent this potential source of difficulty has been to administer a series of

STUDY NOTES BY ANNE | eireandtheclouds
THIS IS NOT FOR SALE. ANY FORMS OF REPRODUCTION IS STRICTLY PROHIBITED.

homogeneous tests, each designed to measure some component of a heterogeneous variable.

- **Kuder-Richardson Formula 20 –** is the statistic of choice for determining the inter-item consistency of dichotomous items, primarily those items that can be scored right or wrong (such as multiple-choice items).
    - Different degrees of difficulty.
    - If test items are more heterogeneous, KR-20 will yield lower reliability estimates than the split-half method.
- **KR21 –** dichotomous items; all the test items have approximately the same degree of difficulty
- **Cronbach's Alpha (α) –** arguably the most widely used and reported measure of internal consistency, especially for tests with items that are scored on a continuous scale (e.g., Likert-scale items, multiple-choice items with partial credit)
    - Is appropriate for use on tests containing nondichotomous items
    - Is the preferred statistic for obtaining an estimate of internal consistency reliability.
    - Is widely used as a measure of reliability, in part because it requires only one administration of the test.
    - Coefficient alpha typically ranges in value from 0 to 1.
    - Streiner (2003b) pointed out, a value of alpha above .90 may be "too high" and indicate redundancy in the items.
- **McDonald's Omega (ω) –** is a measure of internal consistency reliability, similar to Cronbach's Alpha, but often preferred for its ability to handle more complex factor structures; it is a way to assess how well the items on a questionnaire or scale consistently measure a single underlying construct
- **Average Proportional Distance –** a measure used to evaluate internal consistency of a test that focuses on the degree of difference that exist between item scores
    - A measure used to evaluate the internal consistency of a test that focuses on the degree of difference that exists between item scores.

Split-Half Reliability – correlating two pairs of scores obtained from equivalent halves of a single test administered once

- It is a useful measure of reliability when it is impractical or undesirable to assess reliability with two tests or to administer a test twice (because of factors such as time or expense).
- When it comes to calculating split-half reliability coefficients, there's more than one way to split a test—but there are some ways you should never split a test.
- Simply dividing the test in the middle is not recommended because it's likely that this procedure would spuriously raise or lower the reliability coefficient.
- Different amounts of fatigue for the first as opposed to the second part of the test, different amounts of test anxiety, and differences in item difficulty as a function of placement in the test are all factors to consider.
- **Spearman-Brown Formula –** estimate internal consistency reliability from a correlation of two halves of a test
    - It is a specific application of a more general formula to estimate the reliability of a test that is lengthened or shortened by any number of items.
- **Spearman-Brown Prophecy Formula –** estimates how many more items are needed in order to achieve the target reliability

Steps of Spearman-Brown Formula

1. Divide test into equivalent halves
2. Calculate the correlation coefficient (Spearman's rank correlation or Pearson correlation if the data is interval/ratio) between the scores of the two halves
3. Adjust the half-test reliability using Spearman-Brown formula

- **Rulon's Formula (Flanagan-Rulon formula) –** counterpart of the Spearman-Brown formula, a method for assessing the consistency of a test by comparing the scores on two halves of the test

Steps of Rulon's Formula

1. Split the test: Divide the test into two equivalent halves (e.g., odd and even questions)
2. Calculate variances
    - Calculate the variance of the differences between each person's scores on the two halves
    - Calculate the variance of the total scores for each person
3. Apply the formula

- **Odd-Even Reliability –** assign odd-numbered items to one half of the test and even-numbered items to the other half

Interrater Reliability – degree of agreement or consistency

between two or more scorers with regards to particular measure

- **Coefficient of Interrater Reliability –** way of determining the degree of consistency among scorers in the scoring of a test is to calculate a coefficient of correlation
- **Kappa Statistics –** used formula for nominal data
  - **Fleiss Kappa –** agreement between multiple raters (three or more)
  - **Cohen's Kappa –** agreement between two raters
- **Kendal's W –** used for rankings or ordinal data

B | Nature of the Tests

Homogeneity vs. Heterogeneity of Test Items

- **Homogeneous –** test designed to measure one factor (high internal consistency)
- **Heterogeneous –** low internal consistency

Dynamic vs. Static Characteristics

- **Dynamic Characteristics –** a trait, state, or ability presumed to be ever-changing as function of situational and cognitive experience
- **Static Characteristics –** presumed to be relatively unchanging

Restriction vs. Inflation of Range

- Restricted – 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.
- **Inflated –** if the variance of either variable in a correlational analysis is inflated by the sampling procedure, then the resulting correlation coefficient tends to be higher

Speed vs. Power Tests

- **Power Test –** when the time limit is long enough to allow test takers to attempt all items and if some items are so difficult that no test takers can obtain a perfect score
- **Speed Test –** contains items of uniform level of difficulty so that, when given generous time limits, all test takers should be able to complete all the test items correctly

Criterion-Referenced Test – assigned to provide an indication of where a test taker stands with respect to some variable or criterion

- Unlike norm-referenced tests, criterion-referenced tests tend to contain material that has been mastered in hierarchical fashion.

B | Reliability Models and Theories

Classical Test Theory – is the most widely used and accepted model in psychometric literature today.

- Also referred to as the true score (or classical) model of measurement.
- Everyone has a "true score" on a test.
- **True Score –** a value that, according to CTT, genuinely reflects an individual's ability (or trait) level as measured by a particular test.
- Its assumptions allow for its application in most situations

Domain Sampling Theory – seek to estimate the extent to which specific sources of variation under defined conditions are contributing to the test score

- **Domain of Behavior –** the universe of items that could conceivably measure that behavior, can be thought of as a hypothetical construct

Generalizability Theory – a person's test scores vary from testing to testing because of variables in the testing situation

- **Universe –** describe the details of the particular test situation
- **Facets –** the number of items in the test, the amount of training the test scores have had, and the purpose of test administration
- **Universe Score –** the exact same conditions of all the facets in the universe, the exact same test score should be obtained
- **Generalizability Study –** examines how generalizable scores from a particular test are if the test is administered in different situations
- **Decision Study –** developers examine the usefulness of test scores in helping the test user make decisions

Item-Response Theory (Latent-Trait Theory) – provide a theory way to model the probability of that a person with X ability will be able to perform a level of Y

- Stated in terms of personality assessment, it models the probability that a person with X amount of a particular personality trait will exhibit Y amount of that trait on a personality test designed to measure it.
- Latent trait: unobservable
- Manifestation trait: observable
- **Discrimination –** an item differentiates among people with higher or lower levels of the trait, ability, etc., that is being measured
- **Dichotomous Test Items –** only two possible responses
- **Polytomous Test Items –** more than two responses

STUDY NOTES BY ANNE | NOT FOR SALE | ANY FORMS OF REPRODUCTION STRICTLY PROHIBITED

- **Rasch Model –** is a reference to an IRT model with very specific assumptions about the underlying distribution.
- The psychometric advantages of IRT have made this model appealing, especially to commercial and academic test developers and to large-scale test publishers.
- It is a model that in recent years has found increasing application in standardized tests, professional licensing examinations, and questionnaires used in behavioral and social sciences.

Sources of Measurement Error With Typical Reliability Coefficient Used to Estimate Them

Sources of Error	Types of Tests Prone to Each Error Source	Appropriate Measures Used to Estimate Error
Interscorer Differences	Test scored with a degree of subjectivity	Scorer reliability
Time Sampling Error	Tests of relatively stable traits or behaviors	Test-retest reliability (r_{tt}) a.k.a. Stability coefficient
Content Sampling Error	Tests for which consistency of results, as a whole, is desired	Alternate-form reliability (r_{ll}) or split-half reliability
Inter-item Inconsistency	Tests that require inter-item consistency	Split-half reliability or more stringent internal consistency measures, such as KR-20 or coefficient a
Inter-item Inconsistency and Content Heterogeneity Combines	Tests that require inter-item consistency and homogeneity	Internal consistency measures and additional evidence of homogeneity
Time and Content Sampling Error	Tests that require stability and consistency of results, as a whole	Delayed alternate-form reliability

Standard Error of Measurement – provide an estimate of the amount of error inherent in an observed score of measurement

- ➔ **Standard Error of Score –** index of the extent to which one individual's score vary over tests presumed to be parallel
- ➔ **Confidence Interval –** a range or band of the test scores that is likely to contain the true score
  - ◆ Approximately 68% (actually, 68.26%) of the scores would be expected to occur within ±1σ of the true score;
  - ◆ Approximately 95% (actually, 95.44%) of the scores would be expected to occur within ±2σ of the true score;
  - ◆ Approximately 99% (actually, 99.74%) of the scores would be expected to occur within ±3σ of the true score.

Example:

- ➔ Given:
  - ◆ Observed score = 75
  - ◆ SEM = 3
- ➔ For a 68% Confidence Interval:
  - ◆ A 68% confidence interval is typically calculated as ±1 SEM from the observed score.
  - ◆ Lower bound: Observed score - (1 x SEM) = 75 - (1 x 3) = 75 - 3 = 72
  - ◆ Upper bound: Observed score + (1 x SEM) = 75 + (1 x 3) = 75 + 3 = 78
  - ◆ Confidence Interval: 72-78
- ➔ For a 95% Confidence Interval:
  - ◆ A 95% confidence interval is typically calculated as ±1.96 SEM from the observed score (often approximated as ±2 SEM for simplicity). Using the approximation:
  - ◆ Lower bound: Observed score - (2 x SEM) = 75 - (2 x 3) = 75 - 6 = 69
  - ◆ Upper bound: Observed score + (2 x SEM) = 75 + (2 x 3) = 75 + 6 = 81
  - ◆ Confidence Interval: 69-81

- ➔ **Standard Error of the Difference –** statistical measure that can aid a test user in determining a large difference should be before it is considered statistically significant
- ➔ **Standard Error of Estimate –** refers to the standard error of the difference between predicted and observed values

eireandtheclouds@gmail.com | x: rpm_

PSYCHOLOGICAL ASSESSMENT

PmLE | Main References: Cohen & Swerdlik, Kaplan & Sacuzzo, Miller & Lovler, Urbina

VALIDITY

VALIDITY – a test measures what it purports to measure in a particular context

- Is a judgment based on evidence about the appropriateness of inferences drawn from test scores
- **Validity Coefficient –** correlation coefficient that provides a measure of the relationship between test scores and scores on the criterion measure
- **Inference –** is a logical result or deduction
- **Validation –** process of gathering and evaluating evidence about validity
    - **Local Validation Studies –** are absolutely necessary when the test user plans to alter in some ways the format, instructions, language, or content of the test
    - **Ecological Validity –** refers to a judgment regarding how well a test measures what it purports to measure at the time and place that the variable being measured is actually emitted
- Three interrelated ideas that reflect the testing profession's current views about this central and multifaceted concept:
    - The validity of test scores stems from all the accumulated evidence to support their interpretation and uses.
    - As the theoretical understanding of the empirical evidence for test score interpretations accumulates, the validity of inferences (i.e., hypotheses) made on the basis of test scores for various proposed purposes may be enhanced or diminished.
    - Because of the many different purposes to which test scores may be applied, the evidentiary basis for test score interpretations can be derived through a variety of methods.
- Problematic aspects of the traditional view of validity
    - Validity is a property of tests, rather than of test uses or test score interpretations.
    - In order to be valid, test scores should measure some purported construct directly.
    - Score validity is, at least to some extent, a function of the test author's or developer's understanding of whatever construct she or he intends to measure.

Types of Validity

- ➔ **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 individuals with their unique histories and behaviors
- ➔ **Face Validity –** a test appears to measure to the person being related than to what the test actually measures

Summary of Trinitarian View of Validity	
Content Validity	An evaluation of the subjects, topics, or content covered by the item in the test
Criterion-Related Validity	Evaluating the relationship of scores obtained on the test to scores on other tests or measures
Construct Validity	How scores on the test can be understood within some theoretical framework for understanding the construct that the test was designed to measure

A | Content Validity

Content Validity – is a fundamental psychometric property that addresses the extent to which a measurement instrument (e.g., a test, questionnaire, or observational tool) adequately and comprehensively represents all relevant facets or domains of the construct it is designed to measure

- How adequately a test samples behavior representative of the universe of behavior that the test was designed to sample
- *"Does the test content truly reflect the content domain it's supposed to cover?"*
- Achievement tests: is customary to consider a test a content-valid measure

Construct – refers to a theoretical concept, trait, or attribute that cannot be directly observed or measured (e.g., intelligence, depression, anxiety, job satisfaction, mathematical ability)

- ➔ **Construct Underrepresentation –** failure to capture important components of a construct

TOS-BASED REVIEWERS by Anne (rpm_anne, eireandtheclouds@gmail.com) IS ENTIRELY FOR FREE | NOT FOR SALE PLEASE!

➔ **Construct-Irrelevant Variance –** happens when scores are influenced by factors irrelevant to the construct

Core Principles of Content Validity

- **Relevance –** are directly pertinent to the construct being measured
- **Representativeness –** adequately sample all important sub-domains or facets of the construct

Importance of Content Validity

- It ensures that the instrument is indeed measuring what it claims to measure, leading to more accurate and meaningful results
- High content validity enhances the credibility and trustworthiness of the assessment process and its outcomes, particularly for stakeholders such as educators, clinicians, or employers
- While often assessed qualitatively, strong content validity is considered a necessary precursor for establishing other forms of validity, such as construct validity and criterion-related validity
- By systematically covering the entire domain, content validity helps reduce potential biases that might arise from an incomplete or skewed representation of the construct
- For tests used in educational or vocational settings, content validity ensures that the test items reflect the knowledge and skills genuinely required for a given subject area or job

Assessing Content Validity

- Before developing any items, the construct to be measured must be thoroughly and precisely defined; involves identifying all its dimensions, sub-components, and the specific behaviors, knowledge, or skills that comprise it
    - **Test Blueprint –** a detailed plan of the content, organization, and quantity of the items that a test will contain
- **Subject Matter Expert (SME) Review –** panel of qualified subject matter experts (SMEs) independently reviews each item of the instrument against the defined content domain
    - Is the item clearly related to the construct?
    - Does the item contribute to covering the full scope of the construct as defined?
    - Is the item clear, unambiguous, and understandable?
    - Are there any gaps in the content coverage? Are there any redundant items?

Quantitative Indices

➔ **Lawshe's Content Validity Ratio (CVR)**

- For each item, SMEs typically rate its essentiality (e.g., "essential," "useful but not essential," "not necessary")
- The CVR for an item is calculated based on the number of experts who deem the item "essential"
- A positive CVR indicates that more than half of the SMEs consider the item essential
- **Zero CVR –** exactly half of the experts rate the item as essential

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

➔ **Content Validity Index (CVI) –** an be calculated at the *item level (I-CVI)* or *scale level (S-CVI)*

- **I-CVI –** the proportion of SMEs who rate an item as highly relevant (e.g., 3 or 4 on a 4-point relevance scale).
- **S-CVI –** the average of the I-CVIs for all items on the scale (S-CVI/Ave) or the proportion of items that achieve universal agreement (e.g., all experts rate it as 3 or 4) (S-CVI/UA)

- **Pilot Testing and Cognitive Interviewing –** the instrument can be pilot tested with a sample from the target population

B | Criterion-Related Validity

Criterion-Related Validity – demonstrates how well a test or measurement instrument predicts or correlates with a relevant external criterion

- How adequately a test score can be used to infer an individual's most probable standing on some measure of interest

Criterion – is a separate, independent, and often objective measure that serves as a standard against which the test scores are compared

➔ It should be a direct and relevant measure of the construct or outcome of interest

Types of Criterion-Related Validity

- **Concurrent Validity –** is established when the test scores and the criterion measures are collected at approximately the same time
    - It assesses how well the test can estimate an individual's current standing on the

THIS IS NOT FOR SALE | ANY FORMS OF REPRODUCTION | COPYRIGHTED | eireandtheclouds@gmail.com | x: rpm_anne | tiktok: eireandtheclouds

criterion
    - Often used when a new test is developed to replace an existing, well-validated but perhaps more cumbersome, expensive, or time-consuming measure
    - Example: Administering a newly developed, shorter depression questionnaire to a group of patients and simultaneously comparing their scores with a gold-standard clinical diagnosis or a well-established, longer depression inventory
- **Predictive Validity –** is established when the test scores are collected at one point in time, and the criterion measures are collected at a later point in time
    - It assesses how well the test can predict future performance, behavior, or outcomes
    - Essential for tests used in selection, placement, or diagnostic forecasting, where the goal is to forecast future success or likely outcomes
    - Example: Giving a pre-employment aptitude test to job applicants and then, several months later, correlating their test scores with their actual job performance ratings by their supervisors

Incremental Validity – the degree to which an additional predictor explains something about the criterion measure that is not explained by the predictors already in use

Importance of Criterion-Related Validity

- It demonstrates the practical usefulness of a test
- Provides empirical evidence to support decisions based on test scores, such as admission to academic programs, certification for professions, or clinical diagnoses
- A new test with high concurrent validity can offer a more efficient alternative to an existing, more cumbersome assessment method
- Provides accountability for the use of assessment tools by demonstrating that they effectively serve their intended purpose

C | Construct Validity

Construct Validity – is a central concept in psychometrics and refers to the degree to which a test or measurement instrument adequately measures the theoretical construct it purports to measure

- Is a judgment about the appropriateness of inferences drawn from the test scores regarding individual standings on a variable called constructs

Construct – is a theoretical, abstract concept or attribute that is not directly observable but is inferred from observable behaviors or indicators

➔ Examples

- ◆ Intelligence: Inferred from problem-solving abilities, logical reasoning, and learning speed
- ◆ Anxiety: Inferred from physiological symptoms, behavioral avoidance, and cognitive worry
- ◆ Self-esteem: Inferred from self-descriptions, confidence in actions, and resilience to criticism

Core Principles of Construct Validity

- **Clear Conceptualization –** a precise and well-defined theoretical understanding of the construct, including its dimensions, relationships to other constructs, and expected manifestations
- **Operationalization –** translating the abstract construct into concrete, measurable indicators or test items
- **Empirical Evidence –** gathering data that supports or refutes the hypothesized relationships between the test scores and other variables, consistent with the theoretical framework of the construct

Importance of Construct Validity

- Without construct-valid measures, it's impossible to test these theories empirically or draw meaningful conclusions about psychological phenomena
- It ensures that when someone obtains a score on a test (e.g., an "anxiety score"), that score genuinely reflects their level of the intended construct (anxiety) and not some other unrelated factor
- Research findings derived from instruments with high construct validity are more trustworthy and generalizable, contributing to the scientific rigor of a field
- The ongoing process of establishing construct validity can also refine and advance the very theories about the constructs themselves

Evidences of Construct Validity

- **Evidence of Homogeneity –** broader concept refers to the entire network of theoretical relationships between a construct and other constructs, as well as observable behaviors
- **Evidence of changes with age –** if the construct is

STUDY NOTES BY ANNE
NOT FOR SALE.
ANY FORMS OF REPRODUCTION
IS STRICTLY PROHIBITED
eireandtheclouds@gmail.com | x: rpm_anne | tiktok: @eireandtheclouds

expected to change over time (e.g., cognitive ability in children, emotional maturity), the test scores should reflect these expected developmental patterns
    - Longitudinal studies tracking changes in scores over time
- **Evidence of pretest-posttest changes –** the construct is expected to change in response to a specific intervention (e.g., a therapy for depression, a training)
- **Known-Groups Validity (Group Differences) –** can differentiate between groups of individuals who are known, based on external criteria, to differ on the construct being measured
- **Convergent Validity –** scores on the test undergoing constructed validation tend to correlate highly in the predicted direction with scores on older, more established, and already validated test design to measure the same construct
- **Discriminant Validity –** showing a little relationship between test scores and/or other variables with which scores on the test being construct-validated should theoretically be correlated

Multi-trait Multi-method Matrix (MTMM) – it is a method for assessing the **construct validity** of a set of measures in a study; it provides a structured way to evaluate both **convergent validity** and **discriminant validity** simultaneously

➔ Analysis of 4 Types of Correlation Coefficients

- **Monomethod-monotrait Correlations (Reliability Diagonal) –** the correlations of a trait with itself, measured by the same method, represent the reliability of each measure and should be the highest in the entire matrix
- **Heteromethod-monotrait Correlations (Validity Diagonal) –** represent the same trait measured by different methods, significant correlations here provide evidence of convergent validity, meaning that different measures of the same trait converge or agree
- **Monomethod-heterotrait Correlations (Heterotrait-Monomethod Triangles) –** are between different traits measured by same methods—should be low, providing evidence of discriminant validity, meaning the measures can discriminate between different constructs
- **Heteromethod-heterotrait Correlations (Heterotrait-Heteromethod Triangles) –** are between different traits measured by different methods—should be the lowest in the matrix, further supporting discriminant validity

Table 5.3 A Hypothetical Multitrait-Multimethod Matrix (MTMMM)

		Self-Report			Observation			Projective		
Method	Trait	Anx	Aff	Dom	Anx	Aff	Dom	Anx	Aff	Dom
Self-Report	Anx	(.90)								
	Aff	*.45*	(.88)							
	Dom	*.35*	*.38*	(.80)						
Observation	Anx	**.60**	.23	.10	(.92)					
	Aff	.25	**.58**	−.08	*.47*	(.93)				
	Dom	.12	−.12	**.55**	*.30*	*.32*	(.86)			
Projective	Anx	**.56**	.22	.11	**.65**	.40	.31	(.94)		
	Aff	.23	**.57**	.05	.38	**.70**	.29	*.44*	(.89)	
	Dom	.13	−.10	**.53**	.19	.26	**.68**	*.40*	*.44*	(.86)

Factor Analysis – a class of mathematics procedures designed to identify factors or specific variables that are typically attributes, characteristics, or dimensions on which people may differ

- **Data Reduction.** The most common purpose is to simplify complex data by reducing the number of variables, making dataset easier to work and interpret
- **Structure Discovery.** It helps uncover the underlying structure or dimensions within a set of observed variables.
- **Construct Validation.** It is used to develop and refine measurement scales (like personality tests) to ensure that the items truly measure the intended underlying construct (e.g., intelligence, anxiety)
- **Latent Variables.** Provides an estimate of how strongly each observed variable relates to each factor, known as factor loading
    - **Factor Loading –** conveys information about the extent to which the factor determines the test scores
- **Types of Factor Analysis**
    - **Exploratory Factor Analysis –** to explore and discover the underlying factor structure when the researcher has no preconceived idea about how many factors exist or which variables belong to which factor. *"theory-generating"*
    - **Confirmatory Factor Analysis –** to confirm a hypothesized factor structure. *"theory-testing"*

Principal Component Analysis – to reduce data dimensionality and summarize variance with fewer variables

NOTES BY ANNE | NOT FOR SALE | REPRODUCTION PROHIBITED | x: rpm_anne | tiktok: eireandtheclouds

- **Dimensionality Reduction.** The primary goal of PCA is reducing the number of variables (features) in the data set
- **Variance Maximization.** The principal components are chosen to capture the maximum possible variance in the data
- **Orthogonality.** The principal components are orthogonal (perpendicular) to each other, meaning they are uncorrelated
- **Scree Plot.** A plot of the eigenvalues vs. the number of the principal component. It helps in deciding how many components to retain by looking for an "elbow" or point where the explained variance drops sharply (point of inflection)
- **Explained Variance Ratio.** The proportion of the dataset's variance that lies along each principal component

Two Common Traditional Heuristic Methods in EFA and PCA

- ➔ **Kaiser Criterion (K1 Rule) –** retain all factors or components that have an eigenvalue greater than 1.0
  - ◆ Limitation: tends to overestimate the number of factors, especially when the number of variables is large. It is generally not considered the most accurate method
- ➔ **Elbow Method (Scree Test) –** plot the eigenvalues on the Y-axis against the factor number on the X-axis. Retain all factors that appear before the "elbow" (the point where the slope of the curve sharply levels off)
  - ◆ Limitation: highly subjective, as the "elbow" point is often ambiguous or hard to define, leading different researchers to select different numbers of factors from the same plot

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

TEST NORMS

Standardization – process of administering test to a representative sample of test takers for the purpose of establishing norms

- **Sampling –** the process of developing test by selecting the portion of the whole population
  - **Sample –** portion of the whole population

Types of Sampling Method

- ➔ **Stratified Sampling –** sample of people representing different subgroups (strata) of the population
  - ◆ **Stratified-Random Sampling –** every member of the population had the same chance of being included in the sample
- ➔ **Purposive Sampling –** a non-probability sampling technique where the researcher intentionally selects participants based on specific characteristics or criteria relevant to the research question
- ➔ **Incidental Sampling (Convenience Sampling) –** involves selecting participants based on their accessibility and availability to the researcher; used due to budgeting limitations or other constraints

Norms – are test performance data of a particular group of test takers that are designed for use as a reference when evaluating or interpreting individual test scores

- **Norming –** the process of deriving norms
- **Normative Sample –** group of people whose performance on a particular test is analyzed for reference in evaluating the performance of individual test takers

Types of Norms

- ➔ **Percentile Norms –** raw data from a test's standardization sample converted to percentile form
  - ◆ **Percentile –** expression of the percentage of people whose score on test measure falls below a particular raw score
  - ◆ **Percentage Correct –** refer to the number of items that were answered correctly multiplied by 100
- ➔ **Developmental Norms –** norms developed on the basis of any trait, ability, skill, or other characteristic that is developed, deteriorate, or

affected by stage of life
  - Grade Norms – grade-level basis
  - Age Norms – age-equivalent scores

➔ **National Norms –** derived from a normative sample that was nationally representative of the population

➔ **National Anchor Norms –** equivalency table for scores on the two tests, which provides the tool for such comparison

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

➔ **Local Norms –** typically developed by test users; provide normative information with respect to the local population's performance on some test

Norm-Referenced	Criterion-Referenced
Method of evaluation and a way of deriving meaning from test scores by evaluating an individual test taker's scores and comparing it to scores of a group of test takers	Method of evaluation and a way of deriving meaning from test scores by evaluating an individual score with reference to a standard **Criterion –** a standard on which judgment or decision may be based

Fixed Reference Group Scoring System – the distribution of scores obtained on the test from one group of test takers is used as the basis for the calculation of test scores for the future administration of the test

ACCEPTING OR REJECTING INSTRUMENTS AND TOOLS BASED ON PSYCHOMETRIC PROPERTIES

Reliability Index

Reliability Coefficient	interpretation
0.90 and up	Excellent
0.80 - 0.89	Good
0.70 - 0.79	Adequate
Below 0.70	May have limited applicability

Validity Index

Validity Coefficient	interpretation
Above 0.35	Very Beneficial
0.21 - 0.35	Likely to be Useful
0.11 - 0.20	Depends on the Circumstances
Below 0.11	Unlikely to be Useful

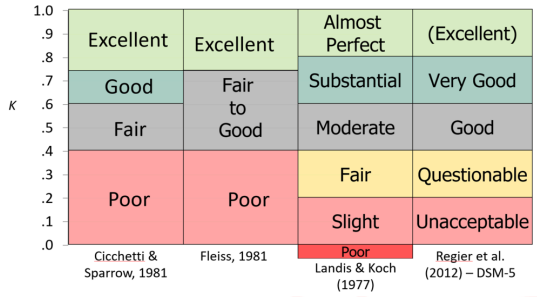
Inter-item Consistency Index

Cronbach's Alpha (α) Value	interpretation
α ≥ 0.90	Excellent
0.80 ≤α< 0.90	Good
0.70 ≤α< 0.80	Acceptable
0.60 ≤α< 0.70	Questionable
0.50 ≤α< 0.60	Poor
α < 0.50	Unacceptable

Interrater Reliability Index

K	Cicchetti & Sparrow, 1981	Fleiss, 1981	Landis & Koch, 1977	Regier et al. 2012 - DSM-5
1.0	Excellent		Almost Perfect	Excellent
0.9				
0.8			Substantial	Very Good
0.7	Good	Fair to Good		
0.6				
0.5	Fair		Moderate	Good
0.4				
0.3	Poor	Poor	Fair	Questionable
0.2				
0.1			Slight	Unacceptable
0.0			Poor	

eireandtheclouds@gmail.com | x: rpm_anne | tik
THIS IS NOT FOR S
ANY FORMS OF REPROD
IS STRICTLY PROHIB



p-Value

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

p-value > 0.10	No evidence against H_o
0.05 < p-value ≤ 0.10	Moderate evidence against H_o
0.01 < p-value ≤ 0.05	Strong evidence against H_o
0.001 < p-value ≤ 0.01	Very strong evidence against H_o
p-value ≤ 0.001	Extremely strong evidence against H_o

RESEARCH METHODS AND STATISTICS (20)

RESEARCH DESIGNS – provide the overarching framework for a study, dictating how data will be collected and analyzed. The following is the most common research design:

- **Experimental Design –** establish robust cause-and-effect relationships between variables by manipulating one or more independent variables and observing their impact on a dependent variable
- **Quasi-Experimental Design –** infer cause-and-effect relationships when true random assignment is not feasible or ethical, such as in real-world educational or clinical settings where intact groups must be used
- **Correlational Design –** quantify the strength and direction of relationships or associations between two or more variables as they naturally occur, without any intervention or manipulation by the researcher
- **Survey Design –** gather data from a sample of individuals, often large, through questionnaires or interviews, in order to describe the characteristics, opinions, or behaviors of a larger population
- **Longitudinal Design –** observe and track changes or developments within the same individuals or groups over an extended period, allowing for the study of developmental trends, stability, and long-term effects
- **Cross-Sectional Design –** gather data from a distinct sample of individuals or groups at a single point in time, providing a snapshot of a phenomenon or population characteristics at that specific moment

STATISTICS

- **Measurement –** the act of assigning numbers or symbols to characteristics of things
- **Error –** collective influence of all the factors on a test score or measurement beyond those specifically measured by the test or measurement
- **Scales –** set of numbers or symbols where properties empirical properties of the objects to which numbers are assigned
  - **Continuous Scale –** countable set of values (can be infinite)
  - **Discrete Scale –** countable in a finite amount of time

Properties of Scales	
Magnitude	The properties of moreness and comparison
Equal Interval	The differences between two variables 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 unit
Ratio	Obtained when nothing of the property being measured exists

4 Levels of Scales of Measurement

- → **Nominal Scale –** categorized using names, labels, or qualities
  - ◆ Gender
  - ◆ Eye color
  - ◆ Types of fruit
  - ◆ Political affiliation
- → **Ordinal Scale –** can be categorized, and the categories have a meaningful order or rank
  - ◆ Educational levels
  - ◆ Satisfaction ratings
  - ◆ Likert scales
  - ◆ Race rankings
- → **Interval Scale –** have all the properties of ordinal data, plus the intervals between consecutive values are equal and meaningful; does not have absolute zero point
  - ◆ Temperature in Celsius or Fahrenheit

eireandtheclouds@gmail.com | x: rpm_

- IQ scores
- Calendar years

→ **Ratio –** have all the properties of interval data, plus a true or absolute zero point
- Height, weight, and length
- Age
- Income
- Number of items sold
- Temperature in Kelvin

A | Descriptive Statistics

Descriptive Statistics – summarize and describe the characteristics of a dataset, providing insights into its distribution, shape, and variability

Frequency Distribution – visualize the spread and patterns of scores

- **Simple FD –** individual scores have been used and the data have not been grouped
- **Grouped FD –** test-score intervals (class intervals) replaced the actual test scores

Frequency Distributions & Graphical Representations

→ **Frequency Distribution Tables –** display the frequency of occurrence of different values or ranges within a dataset

→ **Histograms –** bar charts that show the frequency distribution of numerical data; useful for visualizing the shape, central tendency, and spread of the data

→ **Bar Charts –** used for categorical data to show the frequency or proportion of different categories

→ **Pie Charts –** used to represent parts of a whole for categorical data

→ **Scatter Plots –** used to visualize the relationship between two numerical variables

→ **Line Graphs –** ideal for displaying trends over time

→ **Box Plots (Box-and-Whisker Plots) –** display the median, quartiles, and potential outliers, providing a concise summary of the data's distribution

Measures of Central Tendency – summary of measure that attempts to describe a whole set of data with a single value that represent the middle or center of its distribution

- **Mean –** most stable and useful measure of central tendency

$$\bar{x} = \frac{sum\ of\ the\ items}{number\ of\ items}$$

  - Use for interval and ratio data (normal distribution)
- **Median –** the middle score in the distribution, ordinal, and useful for skewed distribution
- **Mode –** the most frequently occurring score in a distribution of scores
  - Nominal data

Types of Data	Measured Use
Nominal	Mode
Ordinal	Median
Interval/Ratio (skewed)	Median
Interval/Ratio	Mean

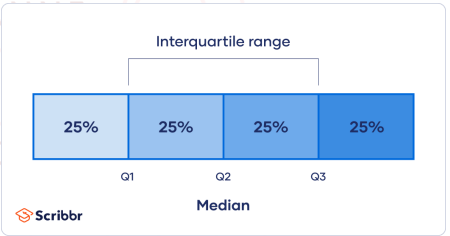
Measures of Variability – summary measure that attempts to describe a whole set of data based how spread out the scores in a distribution

- **Range –** equal to the difference between highest and lowest scores

$$range = highest\ score - lowest\ score$$

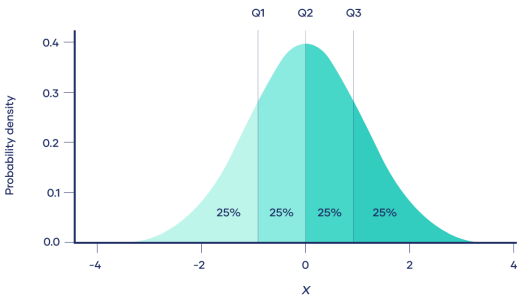
- **Interquartile range –** equal to the difference between Q_3 and Q_1

$$IQR = Q_3 - Q_1$$



- **Semi-interquartile Range –** equal to the interquartile divided by 2

$$SIQR = \frac{Q_3 - Q_1}{2}$$



- **Standard Deviation –** equal to the square root of the average squared deviation about the mean
  - **Variance –** equal to the average of the square of the difference between the scores in the distribution and the mean

PSYCHOLOGICAL ASSESSMENT

PmLE | **Main References: Cohen & Swerdlik, Kaplan & Sacuzzo, Miller & Lovler, Urbina**

Measure of Location – summary of measure that attempts to describe a whole set of data on its location in the distribution

- **Percentile –** used to display position or rank

$$percentile = \frac{\#of\ students\ beaten}{total\ \#\ of\ examinees} \times 100$$

- **Percentage –** means of comparing quantities

$$percentage = \frac{\#\ of\ obtained\ scores}{total\ \#\ of\ items} \times 100$$

- **Quartile –** set of values which has three points dividing the data set into four identical parts

$$quartile = (\frac{p}{4}) \times (N + 1)$$

- **Decile –** a quartile that groups the data into 10 equal parts

$$decile = (\frac{d}{10}) \times (N + 1)$$

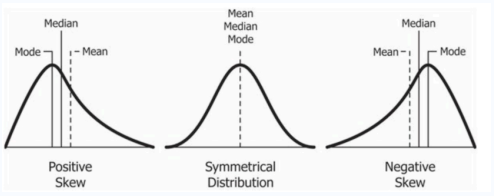
Skewness – the degree of asymmetry observed in a probability distribution; imbalance in the distribution

- Univariate: -3 to +3

Types of Skewness

➔ **Positive Skew –** relatively few of the scores fall at the high end of the distribution (difficult test)
  - ◆ mean>median>mode

➔ **Negative Skew –** relatively few of the scores fall at the low end of the distribution (easy test)
  - ◆ mean<median<mode



Kurtosis – a measure of the tailedness of a distribution; how often outliers occur

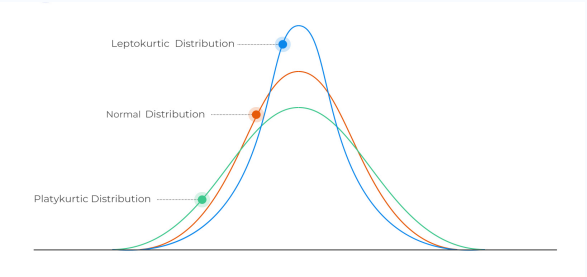
- Univariate: -10 to +10
- The optimal boundary lines for the "upper" and "lower" areas of distribution scores will demarcate the upper and lower 27% of distribution scores
  - 27% if normal
  - 33% if platykurtic

Types of Kurtosis

➔ **Platykurtic –** relatively flat
  - ◆ Thin-tailed, meaning that outliers are infrequent

➔ **Leptokurtic –** relatively peaked
  - ◆ Fat-tailed, meaning that there are a lot of outliers

➔ **Mesokurtic –** somewhere in the middle
  - ◆ Medium-tailed, so outliers are neither highly frequent, nor highly infrequent

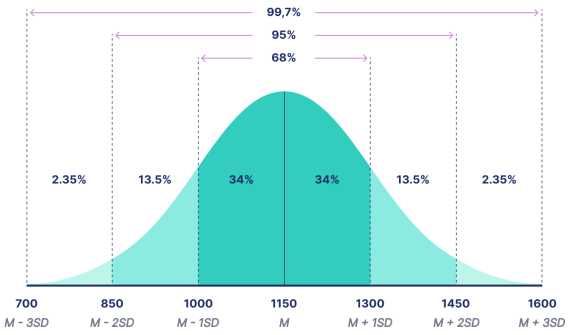


Properties of Kurtosis

	Mesokurtic	Platykurtic	Leptokurtic
Tailedness	Medium	Thin	Fat
Outlier Frequency	Medium	Low	High
Kurtosis	Moderate (3)	Low (<3)	High (>3)
Excess Kurtosis	0	Negative	Positive
Example Distribution	Normal	Uniform	Laplace

Normal Curve

- Bell-shaped, smooth, mathematically defined curve that is highest at its center, perfectly symmetrical yet asymptotic
- Also known as "**Gaussian Curve**"
- Mean = median = mode
- **Tail –** the area on the normal curve between 2 and 3 standard deviation above and below the mean



Area Under the Normal Curve

➔ 50% of the score occur above the mean and 50% of the scores occur below the mean

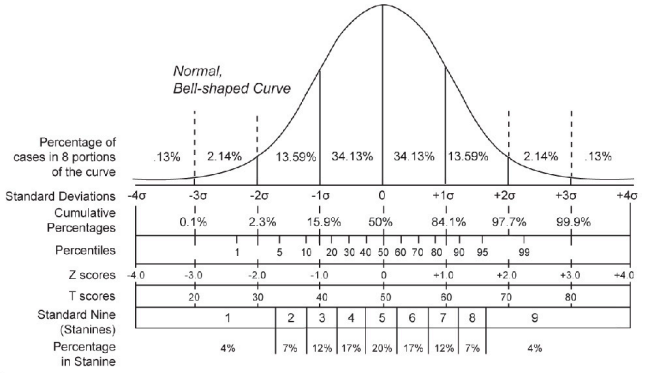
➔ Approximately 34% of all scores occur between

eireandtheclouds@gmail.com | x: rpm_

the mean and 1 standard deviation above the mean
- ➔ Approximately 34% of all scores occur between the mean and 1 standard deviation below the mean
- ➔ Approximately 68% of all scores occur between the mean and
- ➔ ±1 standard deviation
- ➔ Approximately 95% of all scores occur between the mean and ±2 standard deviations

Standard Scores – raw score that has been converted from one scale to another scale

- It tells us the position of test taker's performance relative to other test taker's readily apparent
- **Nonlinear Transformation –** required when the data under consideration are not normally distributed yet needed to compare with normal distribution
- **Linear Transformation –** retains a directed numerical relationship to the original raw score
- **Outlier –** extremely atypical point located at a relatively long distance from the rest of the coordinated points in a scatterplot



Standard Score	Mean	Standard Deviation
Z-Score	0	1
T-Score	50	10
STANINE	5	2
STEN	5.5	2
IQ	100	15
IQ Subtest	10	3
GRE or SAT	500	100

Z-score Formula | Calculating Other Standard Scores

$$Z = \frac{x-\mu}{\sigma} \mid X = Z(SD) + \bar{x}$$

B | Inferential Statistics

Inferential Statistics – uses sample data to make conclusions or predictions about a larger population or dataset

Hypothesis Testing – an educate guess that sets the direction of the of the research; initial response

- **Null Hypothesis (H_0) –** assumes that there is no statistically significant relationship; default
- **Alternative Hypothesis (H_a) –** assumes that there is a statistically significant relationship

Statistical Tests – is method of statistical inferences used to decide whether the data is sufficiently support a particular hypothesis; statistical treatment

- **Parametric Tests –** class of statistical tests that make specific assumptions about the parameters of the population distribution from which the sample data are drawn
- **Nonparametric Tests –** are statistical tests that do not rely on specific assumptions about the shape or parameters of the population distribution; "distribution-free"

Features	Parametric	Non-parametric
Assumed Distribution	Normal distribution (bell curve)	Any distribution (distribution-free)
Assumed Variance	Homogeneous (equal variances across groups)	Heterogeneous (no assumption of equal variance)
Typical Data	Interval or Ratio (Continuous)	Nominal or Ordinal (Categorical/ Ranked)
Data Set Relationship	Independent or Paired (Must meet distributional assumptions)	Any (Flexible with skewed distributions or smaller samples)

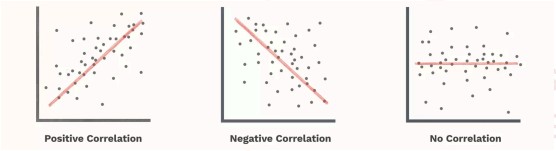
PSYCHOLOGICAL ASSESSMENT

PmLE | **Main References: Cohen & Swerdlik, Kaplan & Sacuzzo, Miller & Lovler, Urbina**

Statistical Tests	Parametric	Non-parametric
Correlation/ Association	Pearson's r	Spearman's Rho
2 Independent Groups	Independent Samples t-test	Mann-Whitney U Test
2 Paired/ Dependents Groups	Paired/Dependent t-test	Wilcoxon Signed-Rank Test
3+ Independent Groups	One-Way ANOVA	Kruskal-Wallis H Test
3+ Repeated Measurements	Repeated Measures ANOVA	Friedman Test

Measure of Bivariate Correlation – connection between 2 variables

- **Correlation –** the degree and direction of correspondence between two things
  - Positive: ↑↑ or ↓↓
  - Negative: ↑↓ or ↓↑
  - No Correlation: —



- **Coefficient of Determination (r^2) –** analyzes how the differences in first variable can be explained by a difference in a second variable
- **Coefficient of Alienation –** proportion of variance in the DV that is not accounted for by the IV
- **True Dichotomy –** naturally divides things into two distinct, mutually exclusive categories, encompassing all possibilities
  - Yes/No
  - True/False
  - Correct/Incorrect
  - Male/Female
- **Artificial Dichotomy –** created by humans, often by assigning categories to continuous variables based on a threshold
  - Passing/Failing
  - High/Low
  - Warning light on/off

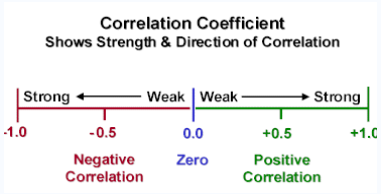
Correlation Tests	Variable Scales	Assumptions
Pearson r	Interval/Ratio + Interval/Ratio	Assumes continuous, normally distributed data with a linear relationship.
Spearman's Rho	Ordinal + Ordinal	Non-parametric; used for monotonic (not strictly linear) relationships.
Biserial	Artificial Dichotomous + Interval/Ratio	The underlying variable is continuous but forcibly split into 2 categories (e.g., Pass/Fail based on score).
Point Biserial	True Dichotomous + Interval/Ratio	Naturally binary variables (e.g., Male/Female, Alive/Dead). A special case of Pearson's r.
Phi Coefficient	True Dichotomous + True or Artificial Dichotomous	Expresses association between two 2x2 nominal categories. Also a special case of Pearson's r.
Tetrachoric	Artificial Dichotomous + Artificial Dichotomous	Used when both underlying variables are actually continuous and normally distributed, but measured as binary.
Kendall's Tau/ Kendall's W	Ordinal/Rank + Ordinal/Rank	Kendall's Tau compares 2 ordinal variables (better for small sample

TOS-BASED REVIEWERS by Anne (rpm_anne, eireandtheclouds@gmail.com) IS ENTIRELY FOR FREE | NOT FOR SALE PLEASE!

		sizes or ties than Spearman). Kendall's W evaluates concordance across 3 or more raters/rankings.
Rank Biserial	Dichotomous Nominal + Ordinal	Measures association between a 2-category nominal variable and an ordinal ranking.

Interpretation of Bivariate Correlation

- ➔ p-Value – significance of the result
  - ◆ 90% CI = 0.01
    - p < 0.01 = significant
    - p > 0.01 = insignificant
  - ◆ 95% CI = 0.05
    - p < 0.05 = significant
    - p > 0.0f = insignificant
- ➔ Algebraic Sign
  - ◆ + = direct relationship
  - ◆ - = inverse relationship
- ➔ Interpretation



Size of Correlation	Interpretation
.90 to 1.00 (−.90 to −1.00)	Very high positive (negative) correlation
.70 to .90 (−.70 to −.90)	High positive (negative) correlation
.50 to .70 (−.50 to −.70)	Moderate positive (negative) correlation
.30 to .50 (−.30 to −.50)	Low positive (negative) correlation
.00 to .30 (.00 to −.30)	negligible correlation

Chi-Squared Tests – primarily used to examine whether two categorical variables are independent in influencing the test statistic; assumption tests; mainly for nominal data

- **Goodness of Fit –** determines if the observed frequencies of a single categorical variable match an expected distribution
- **Test Independence –** is used to determine if two categorical variables are related

Important Assumptions for Both Tests

- ➔ Categorical/Count Data: Observations must be raw counts (frequencies), not percentages, means, or scaled ratios.
- ➔ Independence of Observations: Each subject or event can contribute to only one cell in the table.
- ➔ Adequate Sample Size: The expected frequency in each cell should generally be at least 5 ($E_i \geq 5$). If cells have smaller expected counts, consider using Fisher's Exact Test for 2×2 tables.

Feature	Chi-Square Goodness-of-Fit Test	Chi-Square Test of Independence
Primary Purpose	Evaluates if an observed frequency distribution fits a theoretical or expected distribution	Evaluates whether two categorical variables are associated/ related or independent.
What is Measured	Difference (Observed counts vs. Expected counts).	Relationship & Difference (Association between variables across categories).
Types of Data	Nominal/ Categorical	Nominal/ Categorical
Number of Variables	1 Categorical Variable (with 2 or more levels/ categories).	2 Categorical Variables (each with 2 or more categories forming a cross-tabulation /contingency table).
Degree of Freedom (df)	$df = k - 1$ Where k is the number of categories	$df = (r - 1)(c - 1)$ Where r = rows, c = columns

Comparison of 2 Groups

- **Independent Groups (2 Unrelated Samples)**
  - Parametric: Independent Samples t-Test
    - Example: Comparing test scores between Group A (Tutored) and

Group B (Untutored)

- Non-Parametric Equivalent: Mann-Whitney U Test
  - When to use: Same scenario, but test scores are severely skewed or measured on an ordinal scale (e.g., Likert ratings).

Features	Independent Samples t-Test	Mann-Whitney U Test
Test Type	Parametric	Non-Parametric
Sample Design	2 Separate/ Independent groups	2 Separate/ Independent Groups
Data Requirements	Continuous (Interval/Ratio)	Continuous or Ordinal
Primary Assumption	Normal distribution & equal variances	No distributional assumption (Rank-based)
What it Compares	Difference between Means (μ_1-μ_2)	Difference in Ranks/Medians

- **Dependent Groups (2 Related/Paired Samples)**
  - Parametric: Dependent (Paired) t-Test
    - Example: Comparing student stress levels Before and After an exam on the same group of students.
  - Non-Parametric Equivalent: Wilcoxon Signed-Rank Test
    - When to use: Same pre/post design, but the difference scores violate normality assumptions or use ranked/ordinal metrics.

Features	Dependent (Paired) t-Test	Wilcoxon Signed-Rank Test
Test Type	Parametric	Non-Parametric
Sample Design	1 Group (Paired/ Repeated measures)	1 Group (Paired/ Repeated measures)
Data Requirements	Continuous (Interval/Ratio)	Continuous or Ordinal
Primary Assumption	Normal distribution of difference scores	No distributional assumption (Rank-based)
What it Compares	Mean of the Difference Scores ($\bar{d}$)	Sum of Positive vs. Negative Ranks

Comparison of 2 or More Groups

- **Match Groups –** measures 1 group with ≥ data sets
- **Unmatched Groups –** measures ≥ 3 groups with different data sets

Parametric Test	Non-Parametric Equivalent	Design/ Setup	What It Evaluates
Repeated Measures ANOVA	Friedman Test	1 group, measured across 3 or more conditions /times	Tests for differences across multiple repeated measures on the same subjects.
One-Way ANOVA	Kruskal-Wallis H Test	3 or more independent groups, tested once	Tests for differences among 3+ unrelated groups means across 1 continuous DV.
Two-Way ANOVA	Scheirer-Ray-Hare Test (or aligned rank transform)	2 or more independent variables (IVs), tested on 1 DV	Evaluates main effects of 2 independent factors and their interaction effect.
ANCOVA (Analysis of Covariance)	Quade's Test (or Rank ANCOVA)	1+ categorical IV(s) + 1 continuous Covariate → 1 continuous DV	Blends ANOVA and regression; controls for a covariate variable to reduce error variance.
Mixed-Design ANOVA (Split-Plot)	Non-parametric Mixed Models (e.g., Brunner-Langer approach)	2 or more groups (between-subjects) measured 2+ times (within-subjects)	Combines independent group comparisons with repeated measurements over time.

eireandtheclouds STUDY NOTES THIS IS NOT FOR SALE ANY FORMS OF REPRODUCTION IS STRICTLY PROHIBITED.

Post Hoc Test – (Latin for "after this") is a statistical analysis performed after an Analysis of Variance (ANOVA) or non-parametric equivalent yields a statistically significant result ($p < .05$).

➔ While an ANOVA determines that a significant difference exists among three or more group means, it cannot identify which specific groups differ. Post-hoc tests evaluate pair-by-pair or complex group combinations while controlling for familywise error rate—the inflated risk of making a Type I error (false positive) when running multiple statistical hypothesis tests simultaneously.

Parametric Post-Hoc Tests (Equal Variances Assumed)

➔ **Tukey's Honestly Significant Difference (HSD) Test –** based on the Studentized Range distribution (q) that calculates a single minimum required mean difference between any two groups to be considered statistically significant.

- To conduct all possible pairwise comparisons across group means while strictly maintaining the overall familywise Type I error rate at $\alpha = .05$.
- When to use: Ideal when evaluating all pairwise comparisons in a one-way or two-way ANOVA.
- Assumptions: Continuous data, normal distribution, equal variances across groups (homogeneity), and equal (or near-equal) group sample sizes ($n_1 = n_2 = \ldots = n_k$).

➔ **Scheffé's Test –** derived directly from the F-distribution that evaluates both simple pairwise comparisons and complex linear contrasts.

- Provides maximum protection against false positives when making complex or unplanned group comparisons (e.g., comparing Group 1 against the combined average of Groups 2 and 3).
- When to use: When you need to test complex multi-group combinations, or when sample sizes across groups are markedly unequal.
- Assumptions: Homogeneity of variance and normal distribution.
- Trade-off: It is the most conservative post-hoc test, making it harder to detect small, simple pairwise differences (lower statistical power).

➔ **Bonferroni Correction –** a method that adjusts the significance threshold (α) by dividing it by the total number of comparisons (k)

$$\alpha_{adjusted} = \frac{\alpha}{k}$$

- To control familywise error rate when making a specific, small set of planned pairwise comparisons.
- When to use: When testing a small subset of pre-planned pairwise comparisons rather than all possible combinations.
- Assumptions: Standard parametric assumptions.
- Trade-off: Extremely simple, but becomes excessively conservative as the number of comparisons increases, rapidly raising the risk of Type II errors (false negatives).

➔ **Šidák (Dunn–Šidák) Test –** a modification of the Bonferroni method that uses an exact probability formula to adjust the significance level

$$\alpha_{adjusted} = 1 - (1 - \alpha)^{1/k}$$

- Purpose: To adjust thresholds for multiple comparisons with slightly less conservatism than Bonferroni.
- When to use: Best for a small-to-moderate set of planned comparisons when independence among tests can be assumed.
- Trade-off: Offers higher statistical power than Bonferroni, though differences are minor for small k.

➔ **Dunnett's Test –** a specialized post-hoc test designed exclusively to compare multiple treatment or experimental groups against a single benchmark or control group.

- Reduces the number of necessary comparisons from $\frac{k(k-1)}{2}$ to $k - 1$, preserving statistical power by skipping comparisons between experimental groups.
- When to use: Clinical trials, drug testing, or experimental setups where every treatment condition is compared only to a control/placebo group.
- Assumptions: Equal variances and normal distributions across groups.

➔ **Gabriel's Procedure & Hochberg's GT2 –** modifications of Tukey's KSD specifically designed to handle unequal sample sizes across groups.

- To maintain accurate Type I error rates when group sizes (n) are not balanced.
- Gabriel's Procedure: Best when group

sample sizes are slightly unequal. Higher statistical power than Hochberg's in mild imbalance.
- Hochberg's GT2: Best when group sample sizes are moderately to severely unequal.

Parametric Post-Hoc Tests (Equal Variances NOT Assumed) – when Levene's Test for Equality of Variances is significant ($p < .05$), standard post-hoc tests (Tukey, Scheffé) become invalid due to heteroscedasticity.

- ➔ **Games-Howell Test –** based on Welch's t-test principles and degrees of freedom, designed for situations where homogeneity of variance is violated.
  - Performs pairwise comparisons without requiring equal variances or equal group sample sizes.
  - When to use: Standard choice for all pairwise comparisons when Levene's test fails ($p < .05$) or when sample sizes differ substantially.
  - Assumptions: Continuous data; does not require equal variances or equal n.
- ➔ **Tamhane's T2 –** a conservative post-hoc test based on Sidak's multiplicative inequality and Welch's t-test.
  - Provides conservative protection against Type I errors under heteroscedasticity.
  - When to use: When group variances are unequal and a more conservative test than Games-Howell is preferred.
- ➔ **Dunnett's T3 & Dunnett's C –** designed for pairwise comparisons with unequal variances using studentized maximum modulus distributions.
  - **Dunnett's T3:** Recommended when sample sizes per group are small ($n < 50$) and variances are unequal.
  - **Dunnett's C:** Preferred for large sample sizes with unequal variances.

Non-Parametric Post-Hoc Tests

- ➔ **Dunn's Test –** a non-parametric post-hoc test that analyzes differences in mean rank sums across groups.
  - Identifies which pairs of groups differ significantly following a significant Kruskal-Wallis H test.
  - When to use: Follow-up to Kruskal-Wallis (3+ independent groups on ordinal or non-normal continuous data).
  - Adjustment: Commonly paired with Bonferroni or Holm-Bonferroni corrections to adjust p-values.
- ➔ **Nemenyi Test –** a rank-based post-hoc test similar to Tukey's HSD for non-parametric data.
  - Evaluates all pairwise rank differences across groups.
  - When to use: Follow-up to Kruskal-Wallis or Friedman tests when group sample sizes are equal.
- ➔ **Wilcoxon Signed-Rank Test (with Bonferroni Adjustment) –** pairwise signed-ranked tests applied sequentially across repeated measures conditions.
  - Follow-up analysis to identify specific conditions that differ following a significant Friedman Test (3+ repeated measures).

Regression Analysis – s a statistical technique used to model and analyze the relationship between a dependent variable (also called outcome, criterion, or response variable) and one or more independent variables (predictors or explanatory variables).

- It provides a mathematical framework to:
  - Predict the probable value of an unknown variable based on known variable(s).
  - Quantify the strength and direction of the relationship between variables.
  - Control for confounding variables when assessing specific effects.

Linear Regression of Y on X	$Y = a + bX$ Y value is unknown; the X value is given
Linear Regression of X on Y	$X = c + dY$ X value is unknown; the Y value is given

Normality Test – is used to determine whether a sample dataset is drawn from a population that follows a normal (Gaussian) distribution.

- Checking for normality is a foundational step in statistical analysis because parametric tests assume that data or residuals are normally distributed.
- General Hypothesis Setup for Normality Tests
  - Null Hypothesis (H_0): The data is normally distributed.
  - Alternative Hypothesis (H_1): The data is not

normally distributed.

$p > 0.05 \Rightarrow Fail\ to\ reject\ H_0 \Rightarrow assume\ normality\ (use\ parametri$
$p \leq 0.05 \Rightarrow Reject\ H_0 \Rightarrow nonnormal\ distribution\ (use\ nonparamet$

Note: Unlike standard hypothesis testing where a significant result (p < .05) is often desired, in normality testing, a non-significant result (p > .05) indicates that the normality assumption has been met.

Shapiro-Wilk Test	A test based on the correlation between the data ranks and the expected normal scores. Best Used For: Small to moderate sample sizes (n < 50 to n = 2000). Strengths: Widely considered the most powerful test for detecting departures from normality in small sample sizes. Limitations: Overly sensitive in very large samples (n > 1000), where even minor, trivial deviations from normality produce p < .05.
Kolmogorov-Smirnov (K-S) Test with Lilliefors Correction	Compares the empirical cumulative distribution function (CDF) of the sample data with the theoretical normal CDF. Best Used For: Larger sample sizes (n > 50). Lilliefors Correction: Standard K-S assumes population mean (μ) and variance (σ^2) are known. When these are estimated from sample data, the Lilliefors-corrected K-S test must be used. Limitations: Low statistical power for small sample sizes (n < 30).
Anderson-Darling Test	A modification of the K-S test that gives more weight to the tails of the distribution. Best Used For: Detecting non-normality caused by extreme values, outliers, or heavy/light tails.
	Strengths: More sensitive than the K-S test across various distribution types.
D'Agostino-Pearson Omnibus Test	Combines the sample's skewness (asymmetry) and kurtosis (tail-heaviness/peakedness) into a single omnibus test statistic (χ^2). Best Used For: Moderate to large sample sizes (n ≥ 20). Strengths: Directly pinpoints whether non-normality is due to skewness, kurtosis, or both.
Jarque-Bera Test	A goodness-of-fit test that measures whether sample skewness and kurtosis match a theoretical normal distribution (skewness = 0, excess kurtosis = 0). Best Used For: Econometrics and large continuous data sets (n > 100). Limitations: Poor performance on small sample sizes.

Homogeneity of Variance – (also known as homoscedasticity) is the statistical assumption that different treatment groups, conditions, or levels of a predictor variable have equal population variances.

- Testing for equal variances is a standard preliminary diagnostic before running parametric statistical procedures such as Independent Samples t-tests, Analysis of Variance (ANOVA), and Linear Regression.

Group Comparison Tests (ANOVA/t-Test Context)

➔ **Levene's Test –** evaluates variance equality across two or more groups by performing a One-Way ANOVA on the absolute deviations of individual data points from their group center.

- **Mean-Based (Original Levene's):** Uses deviations from the group mean. Ideal for symmetric, normally distributed data.
- **Median-Based (Brown-Forsythe Test):** Uses deviations from the group median. Highly recommended when data is

STUDY NOTES
THIS IS NOT FOR SALE
ANY FORMS OF REPRODUCTION
IS STRICTLY PROHIBITED
eireandtheclouds@gmail.com | x: rpm_anne | tiktok: eireandtheclouds

skewed or contains outliers.
    - **Trimmed Mean-Based:** Uses deviations from a 10% trimmed mean. Best for heavy-tailed distributions.
- ➔ **Bartlett's Test –** a modified likelihood ratio test (F or χ^2 statistic) based on individual group sample variances.
    - Strengths: High statistical power when data is strictly normally distributed.
    - Limitations: Extremely sensitive to non-normality. Even slight skewness can trigger a false positive ($p \leq .05$) caused by distributional shape rather than actual variance differences.
- ➔ **Fligner-Killeen Test –** a non-parametric test for homogeneity of variances based on ranks of absolute deviations from group medians.
    - Best Used For: Severely non-normal data, heavy outliers, or ordinal/ranked scale continuous metrics.
    - Strengths: Robust against both non-normality and outliers while maintaining solid power.
- ➔ **Hartley's F_{max} Test –** a simple ratio test comparing the largest group variance to the smallest group variance.
    - Best Used For: Quick, manual rule-of-thumb checks when sample sizes across groups are equal ($n_1 = n_2 = \ldots = n_k$) and data is normally distributed.
    - Rule of Thumb: If $F_{max} < 3$ (or $F_{max} < 4$ in larger samples), ANOVA is generally robust enough to proceed.
- ➔ **Cochran's C Test –** compares the single largest variance to the sum of all group variances
    - Best Used For: Detecting a single outlier group that has a disproportionately high variance relative to other balanced groups.

Regression & Time Series Tests (Homoscedasticity of Residuals)

- ➔ **Breusch-Pagan Test –** regresses squared standardized residuals against the predictor variables (X) to detect linear forms of heteroscedasticity.
    - Best Used For: Linear regression models with continuous predictors.
    - Limitations: Sensitive to departures from normality in model residuals.
- ➔ **White's Test –** an extension of the Breusch-Pagan test that regresses squared residuals on predictor variables, their squared terms (X^2), and cross-products (X_1X_2).
    - Best Used For: Detecting both linear and non-linear forms of heteroscedasticity.
    - Strengths: Does not assume a specific linear functional form for heteroscedasticity.
- ➔ **Goldfeld-Quandt Test –** divides the dataset into two subsets (omitting a middle portion) and computes an F-test ratio of the residual sum of squares between the two subsets.
    - Best Used For: Assessing whether variance increases or decreases systematically along an ordered predictor (e.g., income, time).

TEST DEVELOPMENT

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

- Test Conceptualization
- Test Construction
- Test Tryout
- Item Analysis
- Test Revision

A | Test Conceptualization

Test Conceptualization – process of conceptualizing the construct, items included, and design of the test

- Step 1: describe the purpose and rationale for the test
- Step 2: describe the target population for the test
- Step 3: clearly define the key variables of interests
- Step 4: create item specifications
- Step 5: choose item format
- Step 6: specify administration to consider

Some Preliminary Questions

- What is the test designed to measure?
- What is the objective of the test?
- Is there a need for this test?
- Who will use this test?
- Who will take this test?
- What content will the test cover?
- How will the test be administered?
- What is the ideal format of the test?
- Should more than one form of the test be developed?
- What special training will be required of test users for administering or interpreting the test?
- What types of responses will be required of test

takers?
- Who benefits from the administration of this test?
- Is there any potential for harm as the result of an administration of this test?
- How will meaning be attributed to scores on this test?

Pilot Testing – preliminary research surrounding the creation of a prototype of the test
- ➔ **Norm-referenced –** comparisons typically are insufficient and inappropriate when knowledge of mastery is required
- ➔ **Criterion-referenced –** at least two groups of the test takers, one group known to have mastered the knowledge or skill being measured and another group known not have mastered such knowledge or skill

B Test Construction

Test Construction – stage in the process that entails writing test items (or rewriting or revising existing test items), as well as formatting items, setting scoring rules, and otherwise designing and building a test
- **Item Pool –** the reservoir or well from which items will or will not be drawn for the final version of the test
  - It is usually advisable that the first draft contain approximately twice the number of items that the final version of the test will contain
- **Item Banks –** relatively large and easily accessible collection of test questions
- **Computerized Adaptive Testing (CAT) –** items presented to test takers are based in the part on the test taker's performance on previous items
  - **Floor Effect –** low end of ability, trait, or other measurable attribute
  - **Ceiling Effect –** high end of ability, trait, or other measurable attribute
- **Item Branching –** the 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

Types of Item Format
- **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
  - **Checklist –** subject receives a long list of adjectives and indicate whether each one is characteristic of himself/herself
  - **Guttman Scale –** are arranged sequentially from weaker to stronger expressions of attitudes, beliefs, or feelings being measured
- **Selected-Response Format –** select a response from a set of alternatives

Types of Selected-Response Format
- ➔ **Multiple Choice –** three elements: (1) stem, (2) a correct alternative or option, and (3) several incorrect alternatives or options: distractor or foils (25%)
  - ◆ **Effective Distractors –** was chosen by both high- and low-performing groups, which enhances the consistency of a test result
  - ◆ **Ineffective Distractors –** may hurt the reliability of the test because they are time-consuming to read and can limit the number of good items
  - ◆ **Cute Distractors –** less likely to be chosen, may affect the reliability of the test takers, who may guess from the remaining options
- ➔ **Matching Item –** presented with two columns (premises) on the left (responses) on the right
- ➔ **Binary Choice –** contains only two possible responses (true-false item or forced-choice item) (50%)

- **Constructed-Response Format –** supply or create the correct answer
  - **Completion Item –** provide a word or phrase that completes a sentence
  - **Short-Answer –** word, term, sentence, or paragraph may qualify
  - **Essay –** respond to a question by writing a composition

Scaling – is the process of constructing measured scales and assigning numerical values or scores to descriptors, traits, attributes, or performance levels.
- **Notion of Absolute Scaling –** procedure obtaining a measure of item difficulty across samples of test takers who vary in ability
- **Developmental & Norm-Referenced Scaling –** compare an individual's performance to the typical performance of a specific normative reference group across stages of human growth or

STUDY NOTES THIS IS NOT FOR ANY FORMS OF RE IS STRICTLY PROHIBITED.

educational progression.

- **Age-Based Scaling (Age-Equivalent Scores) –** a scale where a test taker's raw score is converted to the average score obtained by individuals of a specific chronological age in the norm sample.
- **Grade-Based Scaling (Grade-Equivalent Scores) –** matches a student's raw score to the mean raw score earned by students at a specific grade level in the normative group.

- **Standardized Score Scaling**
  - **Stanine Scale (Standard Nine) –** a 9-point standardized scaling system that converts raw continuous distribution scores into single-digit integers from 1 to 9.
- **Dimensionality-Based Scaling Models**
  - **Unidimensional Scaling –** a scaling procedure built on the assumption that test items measure a single underlying continuum, latent construct, or trait at a time.
  - **Multidimensional Scaling (MDS) –** a statistical and psychometric technique that maps relationships, similarities, or distances among items across two or more underlying latent dimensions or spatial coordinates.
- **Comparative Scaling –** respondents evaluate two or more items, stimuli, or options relative to each other. The resulting data is comparative, ordinal or ratio-like in relative terms, and produces ipsative scores (forced distribution where choices are interdependent).
  - **Paired Comparison –** a respondent is presented with a pair of stimuli (e.g., statements, traits, brands) and asked to select one according to a given criterion (e.g., Which trait is more characteristic of you?).
  - **Rank Order –** are presented with a list of items or stimuli simultaneously and asked to rank or order them from highest to lowest based on a specified attribute (e.g., 1st to 5th preference).
  - **Constant Sum –** respondents are given a fixed total sum of units or points (typically 100 points) and instructed to allocate those points across several items to reflect their relative importance, preference, or weight.
  - **Q-Sort Technique –** a psychometric sorting procedure where respondents sort a large set of statements, items, or traits (typically 60 to 90 items) into a series of piles representing a continuum (e.g., from "Most Characteristic" to "Least Characteristic").
  - **Forced-Choice (Ipsative) Scaling –** respondents are presented with groups of statements (often matched for social desirability) and forced to select which is "Most Like Me" and "Least Like Me".
  - **Dollar-Metric Scaling –** respondents assign monetary values or differences to pairs of stimuli to indicate how much more they would pay or sacrifice for one option over another.
- **Non-Comparative Scaling –** respondents evaluate each item or stimulus independently on a standalone metric (e.g., Likert scale, continuous rating scale) without direct comparison to other items. Yields normative scores.
  - **Continuous Rating Scales (Graphic Rating Scales) –** respondents rate an object by placing a mark at any appropriate position along a continuous line stretching between two extreme endpoints.
    - **Visual Analogue Scale –** a continuous line, usually exactly 100mm in length, anchored by extreme verbal descriptors at each end (e.g., "No Pain" to "Worst Possible Pain").
  - **Itemized Rating Scales –** respondents are provided with a scale that has a number or brief categories associated with each response option and are instructed to select the single category that best describes the object being evaluated.
    - **Likert Scale (Summative Scale) –** measures the degree of agreement or disagreement with a series of statements regarding an attitude, object, or trait.
    - **Semantic Differential Scale –** 7-point rating scale bounded at opposite ends by bipolar adjectives or phrases (e.g., Passive ↔ Active, Weak ↔ Strong, Cold ↔ Warm).
    - **Stapel Scale –** a unipolar rating scale ranging from -5 to +5 (usually 10 categories) without a neutral zero point, designed to measure how accurately a single adjective or adjective descriptor

SHARED NOTES BY ANNE
THIS IS NOT FOR SALE.
ANY FORMS OF REPRODUCTION
IS STRICTLY PROHIBITED.

describes an object.
    - **Thurstone Scale (Method of Equal-Appearing Intervals) –** an itemized attitude scaling method where statements are assigned scale values based on expert judges' evaluations, making the physical distance between intervals equal across the scale continuum.

Item Scoring Methods & Test Protocols

- **Dichotomous Scoring –** items are scored in binary fashion (0 or 1).
    - Applications: Cognitive ability tests, multiple-choice exams, True/False inventories.
- **Polytomous Scoring –** items are assigned numerical points across multiple categories (> 2 response options).
    - Applications: Rating scales, partial-credit performance tasks, Likert items (1 to 5).
- **Reverse/Keyed Scoring –** prevents and controls for acquiescence response set bias (the tendency to agree with all items regardless of content) by deliberately writing negatively phrased items.
- **Weighted & Differential Item Scoring**
    - **Unweighted Scoring –** every item contributes equally (1 point per item) to the total score.
    - **Empirical/Differential Weighting –** items are assigned different point weights based on difficulty, factor loading weights, or discrimination index (a_i parameter in IRT).
- **Standard Setting & Cutoff Scoring Methods (Criterion-Referenced)**
    - **Angoff Method –** expert judges estimate the proportion of "minimally competent candidates" who would answer each test item correctly. The sum of these probabilities forms the passing cut-score.
    - **Nedelsky Method –** judges evaluate multiple-choice options and identify how many distractors a minimally competent candidate could eliminate as wrong per item.
    - **Bookmark Method –** test items are ordered by IRT item difficulty. Judges place a "bookmark" at the item where a minimally competent test-taker has a 67% probability (RP_{67}) of answering correctly.
- **Latent Trait Estimation (IRT Theta Scoring)**
    - In modern psychometrics (Item Response Theory), scoring shifts from simple sum scores (X) to estimating a candidate's latent ability parameter (θ):
        - **Maximum Likelihood Estimation (MLE) –** estimates θ by finding the trait level that maximizes the likelihood of observing the candidate's specific response pattern.
        - **Bayesian Estimation (EAP/MAP) –** combines response pattern likelihood with a prior population distribution to estimate latent scores.

C | Test Tryout

Test Tryout – administered to a representative sample of test takers under conditions that stimulate the conditions that the final version of the will be administered under

- The test should be tried out on people who are similar in critical respects to the people for whom the test was designed
- Informal rule of thumb: no fewer than 5 subjects and preferably as many as 10 for each item on the test—the more subjects in the tryout, the better
- **Phantom Factors –** factors that actually are just artifacts of the small sample size
    - A definite risk in using too few subjects during test tryout comes during the factor analysis findings

"Good Item" – reliable and valid, and can answered by the test taker correctly by high scores on the test as a whole

Pseudobulbar Affect (PBA) – a neurological disorder characterized by frequent and involuntary outbursts of laughing or crying that may or may not be appropriate to the situation

D | Item Analysis

Item Analysis – statistical procedure to analyze items and evaluate test items

- Employed to assist in making judgments about which items are good as they are and which items need to be revised, or discarded
- **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

STUDY NOTES BY ANNE | NOT FOR SALE. | ANY FORM OF REPRODUCTION IS STRICTLY PROHIBITED. | x: rpm_anne | tiktok: eireandtheclouds

Guidelines of Writing Items

- ➔ Target Single Latent Construct: Each item must evaluate one, and only one, specific skill, knowledge domain, or psychological attribute.
- ➔ Avoid Double-Barreled Statements: Never combine two distinct ideas or constructs into a single item (e.g., "I feel anxious and depressed at work").
- ➔ Avoid Absolute Qualifiers: Terms like "always," "never," "all," or "none" signal extreme statements that are easily identified as false by test-wise respondents.
- ➔ Minimize Negative Phrasing & Double Negatives: Avoid negative stems (e.g., "Which of the following is NOT untrue?") which increase cognitive load without adding construct validity. If negative words are required, emphasize them (e.g., NOT, EXCEPT).
- ➔ Calibrate Reading Level: Match language complexity to the target population so that reading ability does not act as a confounding variable.

Item-Difficulty Index – obtained by calculating the proportion of the total number of test takers who answered correctly

- Denoted as "*p*"
- The higher the value, the easier the question
- A particular item is too easy or too difficult, the item must be rewritten or discarded

Item Difficulty Range	Level of Difficulty
0.00 - 0.19	Very Difficult
0.20 - 0.39	Difficult
0.40 - 0.60	Average
0.61 - 0.79	Easy
0.80 - 1.00	Very Easy

Item Endorsement Index – statistics provides not a measure of of the percent of people passing the item but a measure of the percent of people who said yes to, agreed with, or otherwise endorsed the item

$$P = \frac{R}{T}$$

- *P* is the item difficulty index
- *R* is the number of students who answered the question correctly
- *T* is the total number of students who attempted the question

Optimal Level of Item Difficulty (Optimal Difficulty) – refers to the ideal level of difficulty for the questions on a test or assessment, designed to maximize the effectiveness and quality of the test

- It is usually the midpoint between 1.00 and the chance of success proportion, defined by the probability of answering correctly by random guessing

$$p = \frac{1+c}{2}$$

- The general principle is that item difficulty should be roughly midway between the chance score and a perfect score.
  - **Chance Score –** the proportion of correct answers expected if test-takers were simply guessing.
    - For a 4-option multiple-choice question, the chance score is 1/4 or 0.25
    - For a 5-option question, it's 1/5 or 0.20
  - **Perfect Score –** 1.0 (or 100% correct)
- The ideal difficulty level for a multiple-choice item is slightly higher than midway between the chance score and 1.0
  - **4-option multiple-choice –** the midpoint between the chance score (0.25) and 1.0 is 0.625. A slightly higher value, around 0.74, is often cited as an optimal difficulty level for a four-response multiple-choice item
  - **5-option multiple-choice –** the midpoint between the chance score (0.20) and 1.0 is 0.60. A difficulty level around 0.70 is often considered optimal

Item-Reliability Index – indication of the internal consistency of a test

- The higher the index, the greater the test's internal consistency

Item Reliability Index	Internal Consistency
0.69 & below	May have limited applicability
0.70 - 0.79	Adequate
0.80 - 0.89	Good
0.90 & above	Excellent

STUDY NOTES THIS IS NOT F ANY FORMS OF RE IS STRICTLY PR eireandtheclouds@gmail.com | x: rpm_anne | eireandtheclouds

Item-Validity Index – designed to provide an indication of the degree to which a test is measuring what it purports to measure

- The greater the item-validity index, the greater the test criterion-related validity

Validity Coefficient	Interpretation
Below 0.11	Unlikely to be useful
0.11 - 0.20	Varies on the circumstances
0.21 - 0.35	Likely to be useful
Above 0.35	Very beneficial

Item-Discrimination Index – indicate how adequately an item separates (discriminate) between high scores and low scores on an entire test

- Denoted as "*d*"
- The higher the value of "d", the greater the number of high scores who answered correctly

$$D = \frac{R_H - R_L}{N}$$

- D is the discrimination index
- R_H is the number of students in the high-scoring group who answered the item correctly
- R_L is the number of students in the low-scoring group who answered the item correctly
- N is the number of students in one of the groups (i.e., the size of the high-scoring group or the low-scoring group, which are equal)

Range	Verbal Description
0.09 - 0.019	Poor Item
0.20 - 0.29	Fair Item
0.30 - 0.39	Good item
0.40 & above	Very Good Item

Item Characteristic Curves – graphic representation of item difficulty and discrimination

- The steeper the shape, the greater the item discrimination

Other Considerations in Item Analysis

- **Guessing –** test takers may be instructed to provide an answer only when they are certain (no guessing) or to complete all items and guess when in doubt
- **Item Fairness –** the degree, if any, a test item is biased
    - **Biased Test Items –** an item that favors one particular group of examiners in relation to another when differences in group ability are controlled
    - **Speed Test –** item analyses of tests taken under speed conditions yield misleading or uninterpretable results. The closer an item is to end of test, the more difficult it may appear to be
- **Qualitative Item Analysis –** general term for various nonstatistical procedures designed to explore how individual test items work
    - **"Think Aloud" Test Administration –** designed to shed light on the test taker's thought processes during the administration of test
    - **Extra Panels –** panel members met with the publisher to obtain an understanding of the history and philosophy of test battery and to discuss and define the problem of bias

E | Test Revision

Test Revision – modify a test's content or format for the purpose of improving the test's effectiveness as a tool of measurement

- Characterized each item according to its strengths and weaknesses
- **Cross-Validation –** the revalidation of a test on a sample of test takers other than those on whom test performance was originally found to be valid predictor of some criterion
    - **Validity Shrinkage –** the decrease in item validities that inevitably occurs after cross-validation of findings
- **Co-Validation (Co-Norming) –** a test validation process conducted on two or more test using the same sample of test takers
- Quality assurance during test revision
    - **Training –** form of written guidelines for test administration
    - **Anchor Protocol –** a test protocol scored by a highly authoritative scorer that is designed as a model for scoring and a mechanism for resolving scoring discrepancies
    - **Scoring Drift –** a discrepancy between scoring in an anchor protocol and the scoring of another protocol
- Evaluating the properties of existing tests and guiding test revisions
- Determining measurement equivalence across

eireandtheclouds STUDY NOTES BY ANNE THIS IS NOT FOR SALE ANY FORMS OF REPRODUCTION IS STRICTLY PROHIBITED eireandtheclouds@gmail.com | x: rpm_anne | tiktok: eireandtheclouds

test-taker populations

- **Differential Item Function (DIF) –** wherein an item functions differently in one group of test takers as compared to another group of test takers known have the same level of the underlying trait
    - **DIF Analysis –** test developers scrutinize group-by-group item response curves
    - **DIF Items –** items that respondents from different groups at the same level of the function of their group membership

USES, BENEFITS, LIMITATIONS OF ASSESSMENT TOOLS AND INSTRUMENTS (13)

CONTEXTS OF REFERRAL

- **Psychiatric Context –** often involve assessing individuals suspected of having psychological disorders, such as mood disorders, anxiety disorders, or psychotic disorders
    - **Psychiatrists –** trained medical doctors that administer drugs and medications
    - **Psychologists –** guided by the biopsychosocial model in explaining the causation of the disorder
- **General Medical Context –** includes referrals from physicians or other medical professionals for patients with medical conditions that may have psychological components or for whom psychological factors may be impacting their physical health
    - Examples include assessing cognitive impairments following a stroke or evaluating a patient's adjustment to a chronic illness
- **Legal Context –** are often court-ordered or requested by attorneys, may be used to determine competency to stand trial, assess for psychological damages in personal injury cases, or evaluate parenting capacity in custody disputes
    - Example: child custody, annulment of marriage
- **Educational Context –** are conducted to identify learning disabilities, ADHD, or other learning-related difficulties that may be impacting a student's academic performance and can also be used to determine appropriate educational accommodations or to assess for giftedness
- **Industrial Context –** assessments may be used for personnel selection, placement, training, or performance evaluation and can help organizations identify the best candidates for specific roles or assess the effectiveness of training programs
- **Psychological Clinic Context –** includes referrals from individuals seeking help for personal or interpersonal problems, allows for a broader range of assessment purposes, including diagnostic clarification, treatment planning, and personal growth

ASSESSMENT TECHNIQUES

- **Traditional Assessment –** the conventional methods of evaluating student learning, typically involving pen-and-paper tests or other in-person activities
    - Pros: easy to administer and score, widely accepted, and can be used to assess a wide range of knowledge and skills
    - Cons: May not accurately reflect real-world application of knowledge, can be susceptible to test anxiety, and may not cater to diverse learning styles
- **Tele-Assessment –** is the use of technology to conduct assessments remotely, often utilizing online platforms, video conferencing, or other digital tools
    - Pros: can be more flexible and convenient, can be adapted for diverse learners, and may offer opportunities for more authentic assessments
    - Cons: requires access to technology and reliable internet, may present challenges with technical issues, and may require careful consideration of ethical and security concerns

ASSESSMENT TOOLS

Interviews (Clinical Interview)

- Primary goal is to assess (case formulation and treatment planning) and help (referral) the client
- The goal is to arrive at a diagnosis, pinpoint areas that must be addressed with psychotherapy, determine danger, and solidify therapeutic contract (build rapport)

Issues Addressed by Structured Interview

➔ **Criterion Variance –** amount of error caused by vague guidelines for exclusion or inclusion in different categories

➔ **Information Variance –** variability in amount and type of information derived from interviews with clients

Common Clinical Interview Procedures	
Intake Interview	Is the initial contact between a client and a clinic or clinical practitioner. Its primary function is preliminary data collection to clarify the client's chief complaint, determine appropriateness for services, and gather background history. • **Chief Complaint (CC):** The client's problem stated in their own words. • **History of Present Illness (HPI):** Onset, severity, course, and nature of symptoms. • **Biopsychosocial History:** Medical history, developmental milestones, substance use, educational/occupational history, and family psychiatric history. • **Administrative Setup:** Informed consent, fee structure, confidentiality limits, and service alignment.
Psychodiagnostic Interview (used by Psychiatrists)	Focuses on differential diagnosis and diagnostic classification. While historically associated primarily with psychiatrists, it is routinely conducted by trained clinical psychologists and psychiatric nurse practitioners. • **DSM/ICD Alignment:** Questions are explicitly framed around diagnostic criteria (e.g., DSM-5-TR or ICD-11 symptom checklists). • **Structured & Semi-Structured Formats:** Often utilizes validated diagnostic schedules such as: SCID-5, M.I.N.I. • **Differential Diagnosis:** Systematic rule-out of organic medical conditions, substance-induced symptoms, and overlapping psychiatric disorders.
Behavioral Interview (used by Psychologists)	Popularized by cognitive-behavioral and functional analytic traditions, the Behavioral Interview seeks to operationalize specific behaviors rather than label symptoms abstractly. It focuses on identifying environmental and internal triggers that maintain specific behaviors. **Functional Analysis (ABC Model)** • **Antecedents (A):** Environmental, situational, or cognitive events occurring immediately prior to the target behavior. • **Behavior (B):** The operationalized target behavior (frequency, duration, intensity, and topography). • **Consequences (C):** Reinforcers or punishments (positive/negative) following the behavior that maintain or decrease its occurrence.
Mental Status Examination	Is a structured cross-sectional assessment of a patient's current psychological functioning at a specific point in time. It is analogous to the physical exam in general medicine.
Suicide Assessment Interview	A specialized assessment conducted to evaluate risk, intent, and lethality regarding self-harm or suicide. It is integrated into intake/emergency assessments or conducted whenever acute safety concerns arise.

PSYCHOLOGICAL ASSESSMENT

PmLE | Main References: Cohen & Swerdlik, Kaplan & Sacuzzo, Miller & Lovler, Urbina

Documents

- External source of verification
- **Case History Profile –** a comprehensive record of a patient's medical and personal background, used to understand their condition and guide treatment decisions.
    - It includes information about the patient's current complaints, past illnesses, family history, lifestyle, and other relevant factors
- **Portfolio –** samples of one's ability and accomplishment
- Informed consent

Behavioral Assessment

- **Behavioral Observation –** internal source of verification
- **SORC Model –** stimulus, organismic valuables, actual response, consequences
- Monitoring the actions of others or oneself by visual or electronic means while recording quantitative and/or qualitative information regarding those actions

Common Behavioral Assessment Procedures	
Behavioral Observation and Rating Scale	Involves directly watching, systematically coding, and recording an individual's actions in either their natural environment (naturalistic observation) or a structured laboratory setting (controlled observation). • **Naturalistic Observation:** Observing behavior in real-world settings (e.g., classroom, home, workplace) to maximize ecological validity. • **Quantification & Sampling Strategies:** Event Recording/ Frequency, Interval Recording, Durational and Latency • **Inter-Rater Reliability:** Ensuring consistency across independent observers, often calculated using Cohen's Kappa.
Self-Monitoring	Requires the client to systematically observe and record their own behaviors, internal cognitive processes, or emotional states in real-time as they occur in daily life. • **Measurement Tools:** Behavioral logs, daily mood journals, thought records (CBT), or digital applications (Ecological Momentary Assessment - EMA). • **Target Variables:** Frequency of compulsive behaviors, panic attacks, dietary intake, automatic negative thoughts, or urge levels. • **The Reactivity Effect:** The psychological phenomenon wherein the mere act of measuring one's own behavior alters that behavior (often causing a decrease in undesirable behaviors or an increase in target adaptive behaviors).
Analogue Studies	Involves evaluating client behavior in controlled, artificial, or simulated environments designed to mirror key aspects of real-world situations. This method allows clinicians to manipulate environmental variables in a safe, controlled setting. • **Behavioral Avoidance Tasks (BAT):** Exposing a client with a phobia to a controlled stimulus (e.g., approaching a caged spider or entering a small room) to measure avoidance distance, heart rate, and distress levels. • **Controlled Stress Scenarios:** Simulating

eireandtheclouds@gmail.com | x: rpm_

	public speaking or problem-solving tasks in front of an evaluator to assess stress tolerance and coping strategies. • **Prototypes and Experimental Rigor:** Provides high internal validity by holding external confounding variables constant while observing the target behavior.
Situational Analysis Measure	Place individuals (frequently adults) in standardized, structured real-world or simulated tasks to observe their problem-solving skills, executive functioning, decision-making, and interpersonal dynamics under pressure. • **Assessment Centers:** Standardized multi-exercise evaluations used in organizational and occupational clinical contexts (e.g., leaderless group discussions, in-tray exercises, conflict negotiation). • **Functional Adaptive Tasks:** Evaluating an adult client's independent living skills (e.g., managing a budget, following a recipe, navigating public transit).
Role Play	Is an interactive, action-oriented assessment method where the client is instructed to act out specific social scenarios, conflict situations, or interpersonal encounters with the clinician or a trained actor.
Psychophysiological Method	Quantifies the relationship between psychological states (e.g., anxiety, trauma, stress) and autonomic/central nervous system physiological activity.

Tests

- Includes psychometric and projective tests
- **Standardization –** rigorous process of establishing a test's psychometric principles and test administration
- **3 Characteristics of Psychological Tests**
  - All good tests representatively sample the behaviors thought to measure an attribute or thought to predict an outcome
  - All good tests include behavior samples that are obtained under standardized conditions
  - All good tests have rules for scoring
- **Parties in Psychological Tests**
  - **Test Developers –** creates the test
  - **Test Publisher –** they publish, market, sell, and control distribution of tests
  - **Test Reviewers –** prepare evaluative critiques based on the technical and practical aspects of the tests
  - **Test User –** wide range of professionals
  - **Testtaker –** client, patient, student
    - **Psychological Autopsy –** a reconstructions of deceased individual's psychological profile on the basis of archival records, artifacts, and previous interviews
  - **Test Sponsor –** institutions of government who contract test developers for various testing services
  - **Society at Large –** society's demand for "some way of organizing or systematizing the many-faceted complexity of individual differences"
- **Functions 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
- **Types of Decisions made through Psychological Tests**
  - **Individual –** test takers use their scores to make decisions about themselves

STUDY NOTES BY ANNE
THIS IS NOT FOR SALE.
ANY FORMS OF REPRODUCTION IS STRICTLY PROHIBITED
eireandtheclouds@gmail.com | x: rpm_anne | tiktok: eireandtheclouds

- **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

A | Classifications of Psychological Tests

Based on Number of Test Takers

- **Individually Administered Tests**
  - Can never be administered by group
  - Usually level C tests
- **Group Administered Tests**
  - Usually involves the use of booklets, answer sheets, pen-and-paper
  - Can be administered individually or by group
  - Often level B or A tests but still varies

Based on Test User's Qualifications

- **Level A** – these tests are generally straightforward to administer and interpret with the manual
  - Example: achievement tests
- **Level B** – require some technical knowledge (psychometrics)
  - Typically administered by a Registered Psychometricians with supervisions
  - Example: grouped intelligence tests, objective personality tests
- **Level C** – requires substantial understanding (highly specialized educational background and training)
  - Registered Psychologists
  - Example: intelligence tests, projective tests

Based on the Variable Measured

- **Ability** – what a client can do
- **Typical Performance** – how a client can perform on a regular basis (day-to-day activity)

Feature	Maximum Performance	Typical Performance
Primary Focus	What the client CAN do (Capacity & Maximal Potential)	What the client TYPICALLY does (Habits, Personality & Behavior)
Testing Conditions	controlled, often timed, evaluated in specific settings	reflective of everyday life ("anytime, anywhere")
Nature of Answers	Right or Wrong (Objective accuracy/ Correctness)	No Right or Wrong (All responses reflect personal preference/ state)
Client Motivation	To score as high as possible; produce as many correct answers as possible	To be as honest, open, and self-aware as possible
Key Risk/Bias	Test anxiety, fatigue, insufficient motivation, speededness	Social desirability bias, faking good/bad, lack of self-insight

B | Ability vs. Typical Performance

Ability Tests

- **Achievement Tests** – measures prior learning or learned information
  - Past-oriented
  - Content validity
- **Intelligence Tests** – measures general mental ability
  - Present-oriented
  - Make use of content and construct validity
  - Verbal and/or nonverbal
  - IQ scores
- **Aptitude Tests** – measures potential in learning and performance
  - Future-oriented
  - Content and criterion-related validity
- **Power Tests** – measures extent or depth of skills
  - Spiral omnibus format and ample time is given to complete the test
- **Speed Tests** – measures client's speed or reaction time
  - Client must complete the test as fast as possible
  - Has uniform or simple levels of difficulty

Typical Performance

- **Psychometric Personality Tests** – focuses on stable traits or global factors
  - Considered the most consistent

STUDY NOTES BY ANNE
NOT FOR SALE
ANY FORMS OF REPRODUCTION STRICTLY PROHIBITED
eireandtheclouds@gmail.com | x: rpm_anne | tiktok: eireandtheclouds

- **Interest Inventories –** measures preferences or inclinations
  - Highly correlated with personality tests
  - Stable in some context
- **Values Inventories –** measures things considered as important
  - Inherent values: actual value
  - Attached values: attached importance
- **Attitude Inventories –** measures outlook/dispositions on things
  - Consistent on some extent
- **Other Types of Behavioral Checklists**
  - Usually accomplished by primary caregivers
- **Projective Techniques**
  - **Completion –** client is asked to complete a statement
  - **Expressive –** client is asked to express themselves through creative means
  - **Association –** client is asked to say the first thing they think of when presented a stimulus
  - **Construction –** client is asked to make something out of a stimulus
  - **Choice/Ordering –** client is asked to choose from or to put the stimulus in specific order

Test Batteries – selection of tests and assessment procedures typically composed of tests designed to measure different variables but having a common object

- Fixed Batteries
  - Based on a set standard
  - Typically neuropsychological tests
- Flexible Batteries
  - Client-centered or used for the purpose of responding to the client

Neuropsychological Testing

- Non-invasive tool for screening cognitive-physiological issues
- Presence of issue = brain abnormality/damage
- Components of Neuropsychological Tests
  - Cognitive/information processing should be assessed
  - Should sample relative efficiency of both hemispheres of the brain
  - Should sample anterior and posterior regions of cortical function
  - Should determine the presence of specific deficits
  - Should determine acuteness vs. chronicity onset
  - Should locate intact complex functional systems
  - Should assess affect, personality, and behavior
  - Should have added insights/findings

ASSUMPTIONS IN TESTING & ASSESSMENT

Assumption 1: Psychological Traits and States Exists

- **Trait –** any distinguishable, relatively enduring way in which one individual varies from another
  - **Psychological Traits –** intelligence, specific intellectual abilities, cognitive style, adjustment, attitude, sexual orientation and preferences, psychopathology, specific-, and general personality traits
  - Exist only as construct (an informed, scientific concept developed to describe or explain behavior)
  - Overt behavior
  - Relatively enduring (not 100% expected to be manifested)
- **State –** distinguish one person from another person but relatively less enduring
  - Situation-dependent
- One individual varies from another

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

- Test developers has provided test users with a clear operational definition of the construct under study
- Test developers considered the type of item content that would provide insight into it
- Appropriate ways to score the test and interpret the result
- **Cumulative Scoring –** the more test taker responds in a particular direction as keyed by the test manual as correct or consistent with particular trait, the higher the test taker is presumed to be on the targeted ability or trait
- **Domain Sampling –** sample behavior (test items) from all possible behaviors (items) that could conceivably be indicative of a particular construct

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

- The task in some tests mimics the actual behaviors that the test user is attempting to understand
- Obtained sample of behavior is typically used to make predictions about future behavior
  - **Postdict It –** estimate or suppose something which took place in the past; to conjecture something that occurred

REVIEWER NOTES by Anne
THIS IS NOT FOR SALE
ANY FORMS OF REPRODUCTION
IS STRICTLY PROHIBITED
eireandtheclouds@gmail.com | x: rpm_anne | t: @rpm_anne

beforehand
- **Predict –** say or estimate what (a specific thing) will happen in the future or will be consequence of something

Assumption 4: Test and Other Measurement Techniques have Strengths and Weaknesses
- Competent test users understand a great deal about the test they use as well as appreciate the limitations of the test

Assumption 5: Various Sources of Error are Part of Assessment Process
- **Error –** something that is more than expected; a long-standing assumption that factors other than what a test attempts to measure will influence performance on the test
- **Error Variance –** component of a test score that occur unexpectedly
- Potential sources of error variance
  - Assessors
  - Measuring instruments
  - Random error

Assumption 6: Testing and Assessment can be Conducted in a Fair and Unbiased Manner
- Sensitized test developers and users to the societal demand for fair tests used in fair manner
- Source of fairness-related problems: test users who attempt to use a particular tests with people whose background and experience are different from the background and experience from whom the test was intended

Assumption 7: Testing and Assessment Benefit Society
- Need for instruments to diagnose educational difficulties, neuropsychological impairments, and screen for employment recruits

ASSUMPTIONS IN PSYCHOLOGICAL TESTS

Assumption 1: Psychological Tests Measure What They Purport to Measure or Predict What They are Intended to Predict—also called as "test validity"

Assumption 2: An Individual's Behavior, and therefore Test Scores, will typically remain Stable Over Time—also called "test–retest reliability"

Assumption 3: Individuals Understand Test Items the Same Way
- For example, when asked to respond "true" or "false" to a test item such as "I am almost always healthy," we must assume that all test takers interpret "almost always" similarly

Assumption 4: Individuals Will Report Accurately About Themselves
- E.g., about their personalities, about their likes and dislikes

Assumption 5: Individuals Will Report Honestly their Thoughts and Feelings
- Even if people are able to report correctly about themselves, they may choose not to do so
- Sometimes people respond how they think the tester wants them to respond, or they lie so that the outcome benefits them

Assumption 6: The test score an individual receives is equal to his or her true score plus some error, and this error may be attributable to the test itself, the examiner, the examinee, or the environment

STRENGTHS & WEAKNESSES OF ASSESSMENT TOOLS

Tests – measure device or procedure

Pros	Cons
Standardized Objective Efficient Specific information Normative data	Lack of depth Test anxiety/bias Faking/social desirability Over-reliance

Interviews – method of gathering information through direct communication involving reciprocal response

Pros	Cons
Rich, In-depth Information Flexible Observation of Non-verbal Cues Building Rapport Clarification	Subjectivity/Interviewer Bias Lack of Standardization Time-Consuming Social Desirability Reliability and Validity Issues Memory Effects

Portfolio – samples of one's ability and accomplishments

eireandtheclouds@gmail.com | x: rpm_

Pros	Cons
Demonstrates actual skills and work Contextualized information Creativity and innovation Growth overtime Self-reflection	Subjectivity in evaluation Time-consuming to create and review Lack of standardization Authenticity concerns Limited scope Bias in selection

Behavioral Observation – monitoring actions of others or oneself by visual or electronic means while recording quantitative and/or qualitative information regarding those actions

Pros	Cons
Ecological validity Minimizes self-report bias Identifies behavioral patterns Useful for nonverbal individuals	Observer biasa Reactivity Time-consuming Logistical challenges Limited scope Ethical concerns Interrater reliability

Case History Data – records, transcripts, and other accounts in written, pictorial, or other form that preserve archival information, official and informal accounts, and other relevant items and data

Pros	Cons
Holistic and in-depth understanding Contextual information Longitudinal perspective Uncovers unique patterns Foundation for intervention	Time-consuming and resource-intensive Reliance on retrospective data Subjectivity of information Lack of generalizability Potential for bias Confidentiality and privacy Difficult in verification

Role Play – acting an improvised or partially improvised part in a simulated situation

Pros	Cons
Simulates real-life situations Direct observation of skills	Artificiality Performance anxiety Subjectivity in evaluation
Immediate feedback Identifies gaps and skills Reduces "faking good" Engaging and experiential	Time-consuming Logistical challenges Limited scope Validity concerns

Computers as Tools

- **Local Processing –** scoring may be done on-site
- **Central Processing –** conducted at some central location
- **Teleprocessing –** test-related data may be sent to and returned from this central facility by means of phone lines
- **Simple Scoring Report –** mere listing of a score or scores
- **Extended Scoring Report –** more detailed scoring which includes statistical analyses of the test taker's performance
- **Interpretative Report –** distinguished by its inclusion of numerical or narrative interpretive statements in the report
- **Consultative Report –** usually written in language appropriate for communication between assessment professionals, may provide an expert opinion concerning the analysis of the data
- **Integrative Report –** designed to integrate data from sources other than the test itself into the interpretive report
- **CAPA (computer-assisted psychological assessment) –** the assistance computers provide to the test user, not the test taker
- **CAT (Computer Adaptive Testing) –** the computer's ability to tailor the test to the test taker's ability or test-taking pattern

Pros	Cons
Increased efficiency and speed Enhanced standardization and objectivity Greater accessibility and convenience Improved data management and analysis Multimedia capabilities Reduced test anxiety for some	Technical issues and dependence Lack of human interaction and observation Security and confidentiality concerns Ethical considerations and misuse Initial cost and development Potential for test taker stress in adaptive testing

Actuarial Assessment – an approach to evaluation characterized by the application of empirically

STUDY NOTES BY ANNE
THIS IS NOT FOR SALE
OF REPRODUCTION
Y PROHIBITED.

demonstrated statistical rules as determining factor in assessor's judgment and actions

Mechanical Approach – application of computer algorithms together with statistical rules and probabilities to generate new findings and recommendations

SELECTION OF ASSESSMENT METHODS AND TOOLS (19)

PSYCHOLOGICAL TESTS CMO-34-S-2017 ANNEX B

A | Intelligence Tests and Developmental Scales/Aptitude Tests/Achievement Tests

Intelligence – a multifaceted capacity that manifests itself in different ways across lifespan

- The aggregate, or global, capacity of the individual to act purposeful, to think rationally and to deal effectively within their environment

$$IQ = \frac{mental\ age}{chronological\ age} \times 100$$

- **Francis Galton –** intelligent persons were equipped with the best sensory abilities; tests of visual acuity or hearing ability are, in a sense, test of intelligence
- **Alfred Binet –** when one solves a particular problem, the abilities used cannot be separated because they interact to produce the solution: judgment, reasoning, memory, and abstraction
- **David Wechsler –** is the aggregate or global capacity of the individual to act purposefully, to think rationally and to deal effectively with their environment
- **Two-Factor Theory of Intelligence by Charles Spearman**
  - ***g* (general factor)** representing the portion of the variance that all intelligence tests have in common,
  - And the remaining portions of the variance being accounted for either **specific component (s),**
  - or by **error component (*e*)** of this general factor
- **Raymon B. Cattell's Two Major Types of Cognitive Abilities**
  - **Crystallized Intelligence (Gc) –** acquired skills and knowledge that are dependent on exposure to a particular culture as well as on formal and informal education; retrieval of information and general knowledge
  - **Fluid Intelligence (Gf) –** nonverbal, relatively culture-free, and independent of specific instructions
- **Horn proposed the addition of several factors:**
  - **Vulnerable Abilities –** they decline with age and tend not to return to preinjury levels following brain damage
  - **Maintained Abilities –** tend not to decline with age and may return to preinjury levels following brain damage

INDIVIDUAL TEST ADMINISTRATION (VERBAL & NONVERBAL TESTS)

Stanford-Binet Intelligence Scales, Fifth Edition (SB-5)

- **Purpose:** to assess intellectual and cognitive ability in individuals ranging from early childhood to mature adulthood
- **Population:** 2 to 85+ years.
- **Scores and Scoring:**
  - **Full Scale IQ (FSIQ):** The overarching summary measure of general intellectual ability, derived from all 10 subtests.
  - **Abbreviated Battery IQ (ABIQ):** A brief estimate of cognitive functioning based on two subtests (one verbal, one nonverbal), often used for quick screening.
  - **Domain Scores:** Separate composite scores for the Verbal IQ and Nonverbal IQ domains.
  - **Factor Indices:** Individual composite scores for each of the five factors (FR, KN, QR, VS, WM)—each available in Verbal, Nonverbal, and Combined formats.
  - **Normative Standards:**
    - **Standard Scores:** Typically scaled with a Mean of 100 and a Standard Deviation of 15.
    - **Subtest Scaled Scores:** Have a Mean of 10 and a Standard Deviation of 3.
    - **Percentile Ranks and Age Equivalents:** Provided to help translate scores into practical developmental or educational terms.
- **Time:** varies depending on the examinee's age and cognitive level
  - **Full Scale Administration:** Generally takes between 45 to 75 minutes.
  - **Young Children (Ages 2–5):** Often shorter, taking around 30 to 45 minutes due to shorter attention spans.
  - **Older Children and Adults:** Closer to 60 to 75 minutes.

STUDY NOTES BY ANNE. THIS IS NOT FOR SALE. ANY FORMS OF REPRODUCTION IS STRICTLY PROHIBITED.
eireandtheclouds@gmail.com | x: rpm_anne | tiktok: eireandtheclouds

- **Abbreviated Battery (Screening):** Takes significantly less time, usually around 15 to 20 minutes.

- **Scales and Domains Measured:**
  - **Fluid Reasoning (FR):** The ability to solve novel problems by extracting rules, analyzing patterns, and applying logic.
    - **Nonverbal**: Object Series/Matrices (solving visual pattern sequences)
    - **Verbal**: Early Development/Verbal Absurdities (evaluating logical reasoning through language and pictures)
  - **Knowledge (KN):** Crystallized intelligence, representing an accumulated fund of general information acquired through home, school, and environment.
    - **Nonverbal**: Procedure/ Knowledge Identification (identifying objects and actions)
    - **Verbal**: Vocabulary (defining words and demonstrating verbal knowledge)
  - **Quantitative Reasoning (QR)**: The ability to solve numerical problems, understand mathematical concepts, and manipulate numbers.
    - **Nonverbal**: Quantitative/Number Problems (solving math problems using physical objects or pictures)
    - **Verbal**: Verbal Quantitative Reasoning (solving word problems and applying mathematical logic)
  - **Visual-Spatial Processing (VS)**: The ability to perceive, analyze, and synthesize visual patterns and spatial orientations.
    - **Nonverbal**: Form Board/Block Patterns (manipulating physical blocks or matching visual patterns)
    - **Verbal**: Position and Direction (understanding spatial terms and relationships)
  - **Working Memory (WM):** The capacity to store, transform, and manipulate short-term visual and auditory information.
    - **Nonverbal**: Delayed Response/ Block Span (recalling spatial locations or sequences)
    - **Verbal**: Memory for Sentences/Digits (repeating sentences, digits forward and backward)
  - **IQ Qualification**
    - 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

Background & Earlier Versions of SB-5

- **The 1905 Binet-Simon Scale:** In 1904, the French Ministry of Public Instruction commissioned Binet and Simon to construct a practical screening tool to identify children with cognitive delays or intellectual disabilities who required specialized educational placement.
  - Departed from galtonian sensory-motor tests (e.g., reaction time, grip strength) and focused instead on complex mental processes such as judgment, comprehension, attention, and reasoning.
  - Composed of 30 items arranged in ascending order of difficulty.
  - Did not yield a single quantitative IQ score; rather, it provided a qualitative assessment of whether a child was functioning at a normative level for their age.
- **1908 Scale:** Introduced the formal concept of Mental Age (MA). Items were grouped by age levels (ages 3 to 13) based on the age at which a majority of typical children could successfully complete them.
- **1911 Scale:** Refined the age categories and extended the scale into adulthood, laying the foundation for age-based norming in intelligence testing.
- **First Edition:** 1916 Stanford Revision – Lewis M. Terman (Stanford University).
  - Adapted Binet's items for an American population, standardizing the test on 1,000 children and 400 adults.
  - Adapted William Stern's concept of the Intelligence Quotient (IQ) using the ratio formula:

    $ratio\ IQ = \left(\frac{mental\ age}{chronological\ age}\right) \times 100$
  - Established a standardized mean score

of 100 with a standard deviation typically around 15–16.
    - Popularized the name Stanford-Binet Intelligence Scale.

➔ **Second Edition: 1937 Revision** – Lewis M. Terman and Maud A. Merrill.
    - Created two equivalent forms—Form L (Lewis) and Form M (Maud)—to enable retesting without practice bias.
    - Normed on a larger, geographically broader sample (N = 3,184), extending age coverage down to 2 years and up to superior adult levels.
    - Improved task difficulty graduation and standardized instructions, though the sample was criticized for lacking racial diversity (it consisted primarily of native-born white individuals).

➔ **Third Edition: 1960 Revision (Form L-M) & 1972 Re-Norming** – Maud A. Merrill (following Terman's death in 1956); later norming by Robert L. Thorndike (1972).
    - Consolidated the best-performing items from Forms L and M into a single battery known as Form L-M.
    - Replaced the traditional ratio IQ formula with Deviation IQ scores (Mean = 100, SD = 16).
    - First revision to include non-white populations in the normative sample (N = 2,100), correcting historical demographic bias.

➔ **Fourth Edition: 1986 Revision (SB-IV)** – Robert L. Thorndike, Elizabeth Hagen, and Jerome M. Sattler.
    - Replaced Binet's classical age-grouped item format with subtests containing items ordered by difficulty (point-scale format similar to Wechsler scales).
    - Informed by hierarchical models of cognitive abilities (g-factor model).
    - Organized testing into four broad cognitive areas:
        - Verbal Reasoning
        - Abstract/Visual Reasoning
        - Quantitative Reasoning
        - Short-Term Memory
    - Introduced a vocabulary routing test to determine entry levels for subsequent subtests, reducing administration time and examiner fatigue.

➔ **Arrival of the Fifth Edition: SB-5 (2003) –** developed by Gale H. Roid, the Fifth Edition represented a synthesis of the historical strengths of earlier editions and modern psychometric theory.
    - Explicit incorporation of the Cattell-Horn-Carroll (CHC) theory of cognitive abilities.
    - Equal weight given to Nonverbal and Verbal domains across all five factors, allowing accurate evaluation of individuals with language disorders, non-English backgrounds, or hearing impairments.
    - Added items for low-functioning ranges (identifying developmental delays/intellectual disability) and high-end ranges (identifying giftedness).
    - Normed on 4,800 individuals stratified according to the 2001 U.S. Census.

Wechsler Adult Intelligence Scale, Fifth Edition (WAIS-5)

- **Purpose:** to evaluate general intellectual capacity and specific cognitive domains
- **Population:** 16 to 90+ years.
- **Score and Scoring:**
    - **Full Scale IQ (FSIQ):** The primary summary score reflecting general cognitive ability.
    - **Primary Index Scores:** Composite scores for the five cognitive domains (VCI, VSI, FRI, WMI, PSI).
    - **Standardized Norms:**
        - **Index and Full Scale Scores:** Scaled with a Mean of 100 and a Standard Deviation of 15.
        - **Subtest Scaled Scores:** Scaled with a Mean of 10 and a Standard Deviation of 3.
        - **Percentile Ranks and Confidence Intervals:** Provided to capture score precision and relative standing.
- **Time:**
    - **Full Scale IQ (7 Subtests):** Generally takes approximately 45 minutes.
    - **Primary Index Subtests (10 Core Subtests):** Typically takes about 60 minutes.
    - Extended comprehensive evaluations utilizing supplemental subtests may require additional time.
- **Scales and Domains Measured:**
    - **Verbal Comprehension Index (VCI):** Measures verbal concept formation,

NOTES BY ANNE
NOT FOR SALE
REPRODUCTION
PROHIBITED.
x: rpm_anne | tiktok: eireandtheclouds

verbal reasoning, and acquired knowledge acquired through education and environment.
    - Core: Similarities, Vocabulary
    - Supplemental: Information, Comprehension
- **Visual-Spatial Index (VSI):** Evaluates the ability to evaluate visual details, analyze spatial relationships, and construct geometric designs.
    - Core: Block Design, Visual Puzzles
- **Fluid Reasoning Index (FRI):** Assesses the capacity to detect underlying patterns, solve novel abstract problems, and apply logical reasoning.
    - Core: Matrix Reasoning, Figure Weights
    - Supplemental: Arithmetic, Set Relations
- **Working Memory Index (WMI):** Measures the ability to temporarily register, maintain, manipulate, and recall auditory and visual information in conscious awareness.
    - Core: Digit Span, Spatial Span (or equivalent working memory measures)
    - Supplemental: Letter-Number Sequencing, Running Digit Span
- **Processing Speed Index (PSI):** Evaluates the speed and accuracy of scanning, sequencing, and discriminating simple visual information.
    - Core: Symbol Search, Coding
    - Supplemental: Cancellation, Naming Speed

Wechsler Intelligence Scale for Children, Fifth Edition (WISC-V)

- **Purpose**: to evaluate general intellectual capacity and specific cognitive domains to guide clinical, educational, and intervention planning
- **Population**: 6 years 0 months to 16 years 11 months
- **Score and Scoring:**
    - **Full Scale IQ (FSIQ):** The overarching summary score reflecting general cognitive ability across multiple domains (derived from 7 core subtests).
    - **Primary Index Scores:** Composite scores for the five cognitive domains (VCI, VSI, FRI, WMI, PSI).
    - **Ancillary and Complementary Index Scores:** Alternative groupings measuring specific processing capacities (e.g., General Ability Index, Cognitive Proficiency Index, Quantitative Reasoning, Naming Speed).
    - **Standardized Norms:**
        - **Index and Full Scale Scores:** Scaled with a Mean of 100 and a Standard Deviation of 15.
        - **Subtest Scaled Scores:** Scaled with a Mean of 10 and a Standard Deviation of 3.
        - **Percentile Ranks and Confidence Intervals:** Provided to capture score precision and performance relative to same-age peers.
- **Time:**
    - **Full Scale IQ (7 Core Subtests):** Generally takes approximately 45 to 50 minutes.
    - **Primary Index Profile (10 Core Subtests):** Typically takes about 65 to 80 minutes.
    - Extended comprehensive evaluations utilizing secondary or complementary subtests may require additional testing time.
- **Scales and Domains Measured:**
    - **Verbal Comprehension Index (VCI):** Measures verbal concept formation, verbal reasoning, and knowledge acquired through school and environment.
        - Core: Similarities, Vocabulary
        - Secondary: Information, Comprehension
    - **Visual-Spatial Index (VSI):** Evaluates the ability to evaluate visual details, analyze spatial relationships, and construct geometric patterns.
        - Core: Block Design, Visual Puzzles
    - **Fluid Reasoning Index (FRI):** Assesses the capacity to detect underlying conceptual rules, solve novel abstract problems, and apply logical reasoning.
        - Core: Matrix Reasoning, Figure Weights
        - Secondary: Picture Concepts, Arithmetic
    - **Working Memory Index (WMI):** Measures the ability to register, maintain, manipulate, and recall auditory and visual information in conscious awareness.
        - Core: Digit Span, Picture Span
        - Secondary: Letter-Number Sequencing
    - **Processing Speed Index (PSI):** Evaluates the speed and accuracy of visual

Coach Anne's Notes by Anne
THIS IS NOT FOR SALE.
ANY FORMS OF REPRODUCTION IS STRICTLY PROHIBITED.
Like us on Facebook | x: rpm_anne | tiktok: eireandtheclouds

scanning, sequencing, and discriminating simple visual information.
    - Core: Symbol Search, Coding
    - Secondary: Cancellation

Wechsler Preschool and Primary Scale of Intelligence, Fourth Edition (WPPSI-IV)

- **Purpose:** to evaluate general intellectual capacity and specific cognitive domains to guide clinical, educational, and intervention planning for young children
- **Population:** 2 years 6 months to 7 years 7 months (2:6–7:7)
    - Split into distinct developmental age bands (typically ages 2:6–3:11 and ages 4:0–7:7) to align with rapid early childhood development.
- **Time:**
    - **Ages 2:6 to 3:11:** Generally takes approximately 30 to 45 minutes to complete core subtests.
    - **Ages 4:0 to 7:7:** Typically takes about 45 to 60 minutes for core index profiling.
- **Score and Scoring:**
    - **Full Scale IQ (FSIQ):** The overarching summary score reflecting general cognitive ability across core developmental domains.
    - **Primary Index Scores:** Composite scores for the core cognitive domains (VCI, VSI, FRI, WMI, PSI) depending on age.
    - **Ancillary and Secondary Index Scores:** Alternative groupings measuring specialized cognitive capacities (e.g., General Ability Index, Nonverbal Index, Cognitive Competence Index).
    - **Standardized Norms:**
        - **Index and Full Scale Scores:** Scaled with a Mean of 100 and a Standard Deviation of 15.
        - **Subtest Scaled Scores:** Scaled with a Mean of 10 and a Standard Deviation of 3.
        - **Percentile Ranks and Confidence Intervals:** Provided to capture score precision and performance relative to same-age peers.
- **Scales and Domains Measured:**
    - **Verbal Comprehension Index (VCI):** Measures verbal concept formation, verbal reasoning, and word knowledge acquired through environment and early schooling.
    - **Visual-Spatial Index (VSI):** Evaluates the ability to analyze visual details, understand spatial relationships, and construct geometric designs.
    - **Fluid Reasoning Index (FRI):** Assesses the capacity to detect underlying conceptual rules, solve novel abstract problems, and apply logical reasoning (introduced for ages 4 and up).
    - **Working Memory Index (WMI):** Measures the ability to register, maintain, manipulate, and recall auditory and visual information in short-term memory.
    - **Processing Speed Index (PSI):** Evaluates the speed and accuracy of visual scanning, sequencing, and simple decision-making (introduced for ages 4 and up).
- **Subtests:**
    - **Ages 2 Years 6 Months to 3 Years 11 Months**
        - Core Subtests: Receptive Vocabulary, Block Design, Information, Object Assembly, Picture Naming
        - Supplemental/Optional: Picture Memory, Zoo Locations
    - **Ages 4 Years 0 Months to 7 Years 7 Months**
        - Verbal Comprehension Index (VCI): Information, Similarities, Receptive Vocabulary, Picture Naming, Comprehension, Vocabulary, Word Reasoning
        - Visual-Spatial Index (VSI): Block Design, Object Assembly
        - Fluid Reasoning Index (FRI): Matrix Reasoning, Picture Concepts
        - Working Memory Index (WMI): Picture Memory, Bug Search (or equivalent memory tasks)
        - Processing Speed Index (PSI): Bug Search, Animal Coding, Cancellation

GROUP INTELLIGENCE TEST ADMINISTRATION (NONVERBAL TESTS)

Raven's Progressive Matrices (RPM)

- **Purpose:** to evaluate general cognitive functioning and abstract problem-solving capabilities independently of language, reading level, or cultural background
- **Population:**

eireandtheclouds
REVIEWER NOTES BY ANNE
THIS IS NOT FOR SALE.
ANY FORM OF REPRODUCTION IS STRICTLY PROHIBITED.
fb: Anne Gabrielle Cruz | x: rpm_anne | tiktok: eireandtheclouds

- **Colored Progressive Matrices (CPM):** Designed for young children aged 5 to 11 years, older adults, and individuals with cognitive or learning difficulties. Features simpler patterns with colorful backgrounds.
- **Standard Progressive Matrices (SPM):** Designed for children, adolescents, and adults aged 6 to 65+ years of average intelligence. Uses black-and-white geometric patterns.
- **Advanced Progressive Matrices (APM):** Designed for adolescents and adults of above-average intelligence, frequently used for high-ability screening and selection.
- **Standardization:** Normed on extensive international and national representative samples, accounting for age, geographic region, and demographic variations.

- **Score and Scoring:**
  - **Raw Score:** Calculated by awarding 1 point for each correct answer and 0 points for incorrect answers. The maximum raw score on the SPM is 60.
  - **Percentile Ranks:** The primary interpretive metric. Raw scores are converted into age-based percentiles to indicate how an individual's performance compares to same-age peers.
  - **Grades/Classification Levels:** Often categorized into normative bands (e.g., Grade I for Intellectually Superior/At or above the 95th percentile, down to Grade V for intellectually impaired).
  - **Standardized Scores:** Can be converted into scaled scores, IQ equivalents, or stanines (with a mean of 5) depending on the specific edition and manual guidelines.
- **Time:**
  - **Standard Progressive Matrices (SPM):** Typically untimed, but generally completed within 30 to 45 minutes (with a standard cutoff around 45 minutes if structured timing is desired).
  - **Advanced Progressive Matrices (APM):** Usually allotted around 40 to 60 minutes due to the heightened complexity of the items.
  - **Colored Progressive Matrices (CPM):** Usually takes approximately 15 to 30 minutes for young children.
- **Scales and Domains Measured**
  - **Eductive Ability (Fluid Reasoning):** The capacity to make sense of complexity, store and handle abstract information, extract logical rules, and form working hypotheses.
  - **Reproductive Ability:** The facility to recall and reproduce structured information or learned methods when prompted, though fluid problem-solving is the primary focus.
  - **Primary Construct:** It is considered one of the purest psychometric operationalizations of Spearman's general intelligence factor (g).
- **Subtests and Structure:**
  - **Standard Progressive Matrices (SPM) Structure:** Comprises 60 items divided into five sets (A, B, C, D, and E) of 12 items each.
  - **Advanced Progressive Matrices (APM) Structure:** Comprises 48 items divided into two sets (Set I with 12 items for practice/screening, and Set II with 36 items) featuring highly complex, abstract transformations.

Versions of Raven Progressive Matrices

➔ **Colored Progressive Matrices (CPM) –** designed primarily for young children, older adults, and individuals with cognitive or learning impairments, the CPM utilizes colorful visual arrays to maintain attention and reduce cognitive fatigue.
- ◆ **Target Population:** Children aged 5 to 11 years, elderly populations, and clinical populations with developmental, language, or motor delays.
- ◆ **Structure & Format:** Comprises 36 items divided into three sets (Sets A, Ab, and B) of 12 items each. Items are presented with colored backgrounds to make the tasks more visually engaging and accessible.
- ◆ **Key Objective:** To assess the early development of analogical reasoning and fluid intelligence before children can reliably engage with abstract black-and-white matrices, or to evaluate individuals with declining cognitive capacity.
- ◆ **Administration Time:** Typically brief, taking approximately 15 to 30 minutes.

➔ **Standard Progressive Matrices (SPM) –** considered the classic, flagship form of the assessment, the SPM is designed for individuals

of average intelligence across a broad age span, utilizing black-and-white geometric designs.

- **Target Population:** Children, adolescents, and adults aged 6 to 65+ years of average or typical intellectual functioning.
- **Structure & Format:** Comprises 60 items divided into five sets (A, B, C, D, and E) of 12 items each. Items start very simple and systematically increase in complexity, testing rules such as continuous patterns, figure rotation, and complex metric distributions.
- **Key Objective:** To provide a reliable, culture-fair measure of general intelligence (g) and abstract reasoning across a wide demographic spectrum without relying on language or formal schooling.
- **Administration Time:** Generally untimed, but typically completed within 30 to 45 minutes.

➔ **Advanced Progressive Matrices (APM) –** built specifically to assess higher-level intellectual functioning and differentiate individuals at the upper end of the cognitive spectrum, the APM features highly complex, abstract transformations.

- **Target Population:** Adolescents and adults of above-average to superior intelligence, frequently utilized for high-ability screening, academic differentiation, and professional selection.
- **Advanced Progressive Matrices (APM) Structure:** Comprises 48 items divided into two sets (Set I with 12 items for practice/screening, and Set II with 36 items) featuring highly complex, abstract transformations.
- **Key Objective:** To provide rigorous upper-level differentiation where standard tests might suffer from ceiling effects, effectively measuring subtle variations in high fluid reasoning and problem-solving efficiency.
- **Administration Time:** Usually allotted around 40 to 60 minutes due to the demanding complexity of the items.

R.B. Cattell's Culture Fair Intelligence Test (CFIT)

- **Purpose:** to evaluate innate intellectual potential and abstract reasoning ability independently of sociocultural and linguistic conditioning.
- **Population:**
  - **Scale 1:** Designed for young children aged 4 to 8 years and older individuals with developmental, learning, or cognitive impairments. It incorporates both verbal/perceptual and non-verbal tasks.
  - **Scale 2:** Designed for individuals of average intelligence aged 8 to 14 years and for average-ability adults.
  - **Scale 3:** Designed for adolescents and adults of above-average intelligence, high school/college students, and executive or high-ability screening contexts.
- **Time:**
  - **Test 1: Series –** 3 minutes
  - **Test 2: Classification –** 4 minutes
  - **Test 3: Matrices –** 3 minutes
  - **Test 4: Topology** (2.5 minutes)
- **Scales and Domains Measured:**
  - **Fluid Intelligence (G_f):** The fundamental, biologically based capacity to handle abstract information, extract governing rules, and adapt to unfamiliar problem spaces without relying on stored facts or language skills.
  - **Primary Construct:** It serves as a direct operationalization of Cattell's fluid intelligence factor, acting as a reliable proxy for general intelligence (g) free from cultural loading.
- **Subtests and Structure:**
  - **Series (Subtest 1):** Examinees analyze a progressive sequence of abstract figures and select the option that continues the pattern.
  - **Classification (Subtest 2):** Examinees review a set of figures and identify the odd one out that violates the underlying rule.
  - **Matrices (Subtest 3):** Examinees complete a missing cell within a matrix of geometric patterns based on intersecting row and column rules.
  - **Conditions/Topology (Subtest 4):** Examinees identify which alternative figure satisfies a specific spatial rule (e.g., placing a point in a precise geometric relationship relative to boundaries).

Purdue Non-Language Test (PNLT)

- **Purpose:** to evaluate general cognitive capacity and problem-solving potential independently of reading level, language proficiency, or formal schooling

- **Population:** adolescents and adults across diverse occupational, educational, and cultural brackets.
- **Time:**
  - **Working Test Time:** Exactly 25 minutes of focused testing.
  - **Total Administration:** Typically requires roughly 30 to 35 minutes including instructions and distribution.
- **Scores and Scoring:**
  - **Raw Score:** Calculated by summing the total number of items correctly crossed out using the scoring key.
  - **Normative Interpretation:** Raw scores are evaluated against performance bands or converted into percentile ranks/norm-referenced equivalents (such as categorical classifications from Low to High based on manual norms).
- **Domain Measured:**
  - **General Mental Ability (g):** Measures the underlying capacity to reason, perceive abstract relationships, and discern logical rules without relying on language or crystallized knowledge.
  - **Visual-Spatial and Perceptual Discrimination:** Assesses the examinee's ability to analyze geometric figures, process visual variations, and extract governing principles.

Panukat ng Katalinuhang Pilipino (PKP) by Palacio

- **Purpose:** to provide an equitable and culturally grounded evaluation of cognitive functioning for Filipino adolescents and adults
- **Population:** 16 years old and above
- **Time:** 52 minutes
- **Scores and Scoring:**
  - **Raw Scores:** Summed independently for each of the four subtests.
  - **Standard Scores:** Raw scores are converted into standard scores (with a Mean of 50 and a Standard Deviation of 10) to ensure comparability across subtests.
  - **Composite Intelligence Scores:** Generates distinct summary metrics including the Crystallized Intelligence Score (CIS), Fluid Intelligence Score (FIS), and General Intelligence Score (GIS).
  - **Normative Standards:** Mapped onto percentile ranks using normative data gathered across student and professional groups.
- **Scales and Domains Measured:**
  - **Crystallized Intelligence Score (CIS):** Derived from verbal and numerical proficiencies, representing acquired knowledge, language comprehension, and mathematical application.
  - **Fluid Intelligence Score (FIS):** Derived from non-verbal performance, indicating abstract reasoning capabilities and problem-solving efficiency when confronting novel visual stimuli.
  - **General Intelligence Score (GIS):** An overarching composite score calculated by combining the CIS and FIS, offering a comprehensive approximation of overall intellectual functioning.
- **Subtest and Structure:** comprises 135 total items distributed across four distinctly timed subtests utilizing the Filipino language
  - **Vocabulary/Talasalitaan:** Assesses language proficiency and the capacity to comprehend word meanings.
  - **Analogy/Ugnayan:** Evaluates analytical thinking and the ability to perceive logical relationships between concepts.
  - **Numerical Ability/Kakayahan sa mga Bilang:** Measures proficiency in applying basic mathematical concepts and quantitative problem-solving.
  - **Non-Verbal Ability/Isinasalarawang Problema:** Tests abstract reasoning and problem-solving efficiency through the interpretation of visual figures and spatial patterns.

Differential Aptitude Tests (DAT-5)

- **Purpose:** to provide an objective, multidimensional profile of cognitive proficiencies and potential rather than evaluating academic achievement or fixed general intelligence
- **Population:** adolescents (grades 7–12/ages 12–17) and adults
- **Time:**
  - **Full Battery (DAT-5):** Typically requires approximately 2 to 2.5 hours when administered across all subtests.
  - **Individual Subtests:** Most subtests range between 6 to 25 minutes each (e.g., Perceptual Speed takes about 6 minutes, while Reasoning subtests take roughly 20 to 25 minutes each).
  - **DAT Next Generation:** Highly condensed through computerized adaptive testing (CAT), averaging around 12 minutes per subtest (roughly 1 hour total).

STUDY NOTES BY ANNE | NOT FOR SALE. | ANY FORMS OF REPRODUCTION IS STRICTLY PROHIBITED

- **Scores and Scoring:**
  - **Raw Scores:** Calculated by awarding points for correct answers (with standard correction formulas for guessing applied in specific speeded subtests).
  - **Standardized Norms:** Converted into Percentile Ranks, Stanines, STEN scores, or T-scores based on extensive national standardization samples, allowing comparisons against same-age or same-grade peers.
  - **Profile Reporting:** Generates a distinct multi-score aptitude profile highlighting peak strengths and developmental growth areas across all tested subtests rather than a single summary intelligence quotient.
- **Scales and Domains Measured:**
  - **Cognitive Reasoning:** Measures abstract, verbal, and numerical problem-solving capacities.
  - **Spatial and Mechanical Aptitude:** Evaluates visualization of structural forms and comprehension of physical and mechanical principles.
  - **Clerical and Linguistic Efficiency:** Assesses perceptual speed, accuracy, and written language mechanics.
- **Subtests and Structure:**
  - **Verbal Reasoning (VR)/Verbal Analogies:** Evaluates the ability to understand words, perceive relationships between concepts, and think logically using language.
  - **Numerical Ability (NA)/Numerical Calculations & Sequences:** Measures facility in handling basic arithmetic operations, understanding numerical relationships, and solving quantitative problems.
  - **Abstract Reasoning (AR):** Assesses non-verbal reasoning and the ability to perceive patterns, rules, and logical sequences in abstract geometric figures.
  - **Mechanical Reasoning (MR):** Measures the capacity to understand basic mechanical principles, tools, machinery, and physical laws of motion.
  - **Space Relations (SR):** Evaluates the ability to visualize a three-dimensional object from a two-dimensional flat pattern and mentally manipulate shapes in space.
  - **Clerical Speed and Accuracy (PSA):** Measures the speed and precision of response in simple perceptual scanning and comparison tasks.
  - **Spelling (SP):** Assesses an individual's ability to recognize correct spelling in common English words as an element of written language proficiency.
  - **Language Usage (LU):** Evaluates grammar, punctuation, sentence structure, and general facility with written communication.

Thurstone Tests of Mental Alertness

- **Purpose:** to evaluate an individual's general intellectual capacity and cognitive flexibility under strict time constraints.
- **Population:** high school students, college undergraduates, adults, and working professionals across diverse occupational brackets.
- **Time:**
  - **Working Test Time:** Exactly 20 minutes to complete as many of the 126 items as possible.
  - **Total Administration:** Typically takes roughly 25 to 30 minutes including sample items and instructions.
- **Score and Scoring:**
  - **Raw Scores:** Calculated by summing the total number of correct answers across the linguistic items (L) and quantitative items (Q), combined into a Total Score.
  - **Normative Standards:** Raw scores are mapped onto standardized percentile tables relative to specific working populations (e.g., general job applicants, first-line managers, or senior executives) to determine relative standing.
- **Scales and Domains Measured:**
  - **Linguistic Ability (L Score):** Measures verbal comprehension, word knowledge, vocabulary fluency, and the ability to process text-based relationships.
  - **Quantitative Ability (Q Score):** Measures numerical reasoning, mathematical problem-solving, and the capacity to analyze logical series of numbers.
  - **Total Score (L+Q):** An overarching composite score reflecting general cognitive alertness, processing speed, and the efficiency of shifting between verbal and numerical tasks.
- **Subtests and Structure:**
  - **Linguistic Items (72 Total):**
    - **Same-Opposite/Synonyms & Antonyms:** Evaluating word meanings and semantic

STUDY NOTES BY ANNE | NOT FOR SALE. ANY FORMS OF REPRODUCTION IS STRICTLY PROHIBITED.

relationships.
    - **Definitions:** Identifying terms based on descriptive definitions or linguistic properties.
  - **Quantitative Items (54 Total):**
    - **Arithmetic/Word Problems:** Solving quantitative scenarios requiring basic to advanced mathematical operations.
    - **Number Series:** Analyzing sequences of numbers to discover underlying rules, progressions, or logical patterns.
  - **Unique Structure:** Questions alternate between linguistic and quantitative formats, forcing the examinee to constantly "shift gears" and test their cognitive agility.

Wonderlic Cognitive Ability Test (WCAT)

- **Purpose:** to provide employers and human resource professionals with a quick, reliable metric of a candidate's cognitive capacity to learn, adapt, and solve problems on the job
- **Population:** working adults, job applicants, and professionals across broad industrial, corporate, financial, and administrative sectors.
- **Time:** 12 minutes to answer as many of the 50 items as possible.
- **Scores and Scoring:**
  - **Raw Score:** Calculated by awarding 1 point for each correct answer given within the time limit. The maximum possible raw score is 50. No points are deducted for incorrect answers or skipped items.
  - **Wonderlic Score:** The final raw score is often interpreted directly or mapped onto norm-referenced percentile tables relative to specific occupational benchmarks and job families.
- **Scales and Domains Measured**
  - **General Cognitive Ability (g):** Measures overall mental alertness, logical reasoning, and processing efficiency.
  - **Multi-Domain Problem Solving:** Evaluates verbal comprehension, mathematical reasoning, spatial logic, and deductive/inductive analysis within a single composite score.

Watson Glaser Critical Thinking Tests (WGCT)

- **Purpose:** to evaluate how effectively an individual thinks critically when confronted with complex problems, arguments, and evidence
- **Population:** adolescents, university students, working professionals, managers, and executive candidates.
- **Time:** 30 to 40 minutes depending on the specific form (e.g., standard 40-item or 50-item versions).
- **Scales and Domains Measured:**
  - **Recognize Assumptions (R):** The ability to separate statements of fact from unstated assumptions, value judgments, or presuppositions in a given scenario.
  - **Evaluate Arguments (E):** The capacity to judge whether assertions or arguments are strong and relevant versus weak, irrelevant, or logically flawed.
  - **Draw Conclusions (D):** The skill of weighing evidence, drawing warranted inferences, and determining whether logical conclusions logically follow from available data without overgeneralizing.
- **Subtests and Structure**
  - **Inference:** Judging the truth levels of inferences drawn from specific factual statements (ranging from "True" or "Probably True" to "False," "Probably False," or "Insufficient Data").
  - **Recognition of Assumptions:** Identifying unstated premises or presuppositions taken for granted in specific assertions.
  - **Deduction:** Determining whether logical conclusions necessarily follow from given premises or statements.
  - **Interpretation:** Weighing evidence and deciding whether generalizations or conclusions are fully warranted by the data provided.
  - **Evaluation of Arguments:** Distinguishing between strong, pertinent arguments and weak, irrelevant, or superficial arguments concerning controversial questions.
- **Scores and Scoring:**
  - **Raw Score:** Calculated by summing all correct answers across the subtests. No points are subtracted for incorrect responses.
  - **Composite Critical Thinking Score:** The overarching total score reflecting general critical thinking proficiency.
  - **Normative Standards:** Raw scores are mapped onto national and international percentile norms relative to specific reference groups (e.g., job applicants, managers, college graduates) to determine relative standing.

Flanagan Industrial Tests (FIT)

STUDY NOTES BY ANNE
THIS IS NOT FOR SALE.
ANY FORMS OF REPRODUCTION
STRICTLY PROHIBITED

- **Purpose:** to provide a flexible, job-element approach to personnel selection, placement, and vocational guidance
- **Population:** adults, job applicants, current industrial employees, and students undergoing vocational evaluation.
- **Scales and Domains Measured:**
  - **Cognitive and Academic Aptitudes:** Numerical facility, verbal expression, logical reasoning, and memory.
  - **Technical and Perceptual Capacities:** Mechanical comprehension, electrical principles, blueprint/schematic component identification, scale/table reading, and inspection accuracy.
  - **Spatial and Psychomotor Skills:** Manual/structural assembly, pattern reproduction, spatial coordination, and precision handling.
- **Subtests and Structure:**
  - **Arithmetic:** Working quickly and accurately with basic numbers and calculations.
  - **Assembly:** Skill related to visualizing the results of structural and mechanical changes.
  - **Components:** Ability to locate, match, and identify important parts of a whole system or diagram.
  - **Coordination:** Ability to coordinate hand and arm movements smoothly and accurately.
  - **Electronics:** Understanding electrical/electronic principles and analyzing circuit diagrams.
  - **Expression:** Ability to convey ideas clearly in writing and speaking.
  - **Ingenuity:** Ability to devise creative procedures and problem-solve novel mechanical/practical dilemmas.
  - **Inspection:** Ability to spot flaws, blemishes, or discrepancies quickly and accurately.
  - **Judgment and Comprehension:** Ability to read with understanding, reason logically, and use good judgment in interpreting factual materials.
  - **Mathematics and Reasoning:** Understanding basic mathematical concepts and translating ideas into mathematical notations.
  - **Mechanics:** Understanding mechanical principles and analyzing physical movements and forces.
  - **Memory:** Ability to learn and recall unfamiliar terms or paired associates.
  - **Patterns:** Ability to perceive and reproduce simple pattern outlines precisely and accurately.
  - **Planning:** Ability to plan, organize, and schedule operations or work tasks efficiently.
  - **Precision:** Ability to perform delicate work involving small objects and fine motor control.
  - **Scales:** Ability to read graphs, dials, and measurement scales accurately.
  - **Tables:** Ability to read and interpret complex tables quickly and accurately.
  - **Vocabulary:** Evaluating word knowledge and verbal comprehension breadth.
- **Scores and Scoring:**
  - **Raw Scores:** Calculated by summing correct answers minus adjustments for guessing on specific subtests.
  - **Standardized/Occupational Norms:** Raw scores are typically converted into Percentile Ranks or Occupational Stanines relative to specific working populations or job families, allowing employers to determine an applicant's probability of success in a specific role.
- **Administration Time:**
  - **Per Subtest:** Each of the 18 tests is extremely short, typically requiring between 5 to 15 minutes of working time.
  - **Full Battery:** Administering all 18 modules is rare; customized selections typically require anywhere from 1.5 to 3 hours depending on the specific combination chosen for a target job profile.

Philippine Aptitude Classification Test (PACT)

- **Purpose:** to provide guidance counselors, human resource professionals, and psychologists with an empirical tool for educational placement and career counseling
- **Population:** high school students (most optimal around Grade 9/Third Year High School) up to first-year college students, as well as adult job applicants.
  - Rationale for Timing: Administering the test during junior/senior high school allows guidance counselors adequate time for individual score interpretation and proper vocational counseling before graduation.
- **Scales and Domains Measured:**
  - **Perceptual Speed and Acuity:** Speed in visual comparison, scanning,

eireandtheclouds
STUDY NOTES BY ANNE
THIS IS NOT FOR SALE.
ANY FORMS OF REPRODUCTION IS STRICTLY PROHIBITED.

proofreading, and detecting fine details.
    - Verbal Aptitude (English and Filipino): Comprehension of word meanings, analogies, and language structures in both bilingual settings.
    - **General and Quantitative Reasoning:** Facility with numerical operations, number series, and solving broad mathematical/logical reasoning problems.
    - **Spatial and Perceptual Closure:** Ability to retain visual configurations, identify hidden shapes amidst distractions, and mentally visualize 3D objects from 2D representations.
    - **Mechanical and Visualization Skills:** Capacity to understand mechanical motion, geometric patterns, and structural assembly.
- **Subtests and Structure:**
    - **Part I: Speed Tests** (Focuses on rapid perceptual processing and accuracy)
        - Matching Letters/Numbers (Perceptual Speed)
        - Form Matching (Perceptual Speed)
    - **Part II: Power Tests** (Focuses on reasoning level and cognitive depth)
        - Vocabulary & Analogies (Verbal English)
        - Numeric & Number Series/Figural Reasoning (General Reasoning)
        - Paper Form Board & Hidden Figure (Flexibility of Closure)
        - 6. Talasalitaan & Mga Salitang Magkaugnay (Verbal Filipino)
        - Hidden Blocks I & II (Spatial Closure)
        - Patterns, Mechanical Motion, & Assembly (Visualization)
        - Figure Series & Proofreading (Perceptual Acuity)
- **Scores and Scoring:**
    - **Raw Scores:** Calculated per subtest based on correct answers (with standard handling of speeded versus power items).
    - **Qualitative Indices (QI):** Scores are translated into categorical probabilities (e.g., Very High, High, Average, Low, Very Low) indicating the likelihood of success in specific post-secondary programs.
    - **Aptitude Profile:** Generates a multi-dimensional profile mapping strengths across linguistic, quantitative, spatial, and clerical dimensions.
- **Administration Time:**
    - **Part I (Speed Tests):** Strictly timed, typically lasting exactly 4 minutes.
    - **Part II (Power Tests):** Given approximately 1 hour and 45 minutes.
    - **Total Administration:** Roughly 2 hours, making it efficient for group school or corporate settings.

B | Personality Inventory Tests/Projective Tests

Basic Personality Inventory

- **Purpose:** to provide a concise, psychometrically sound profile of personality functioning and psychological distress.
- **Population:** adolescents and adults aged 12 years and older.
    - **Reading Level:** Accessible at approximately a 5th-grade reading level, ensuring broad applicability across varied educational backgrounds.
    - **Populations:** Extensively utilized in clinical, psychiatric, counseling, correctional, and normal normative environments.
- **Scales and Domains Measured:**
    - **Measures of Mood and Personal/Emotional Adjustment:**
        - **Hypochondriasis:** Preoccupation with bodily health, physical complaints, and illness.
        - **Anxiety:** Presence of cognitive and affective symptoms of tension, worry, and emotional reactivity.
        - **Depression**: Feelings of despondency, hopelessness, guilt, and low self-worth.
    - **Measures of Inadequate or Deviant Socialization and Impulse:**
        - **Interpersonal Problems:** Difficulties in maintaining positive relationships, hostility, and friction with others.
        - **Alienation:** Asocial attitudes, feelings of isolation, and personalized standards of right and wrong.
        - **Impulse Expression:** Difficulty tolerating boredom, impatience, risk-taking, and acting on urges without foresight.
    - **Measures of Cognitive Functioning:**

eireandtheclouds
STUDY NOTES BY ANNE
NOT FOR SALE.
ANY FORMS OF REPRODUCTION IS STRICTLY PROHIBITED.

        - **Thinking Disorder:** Cognitive disorganization, confusion, distractibility, and loose associations.
        - **Persecutory Ideas:** Suspiciousness, distrust, and beliefs that others are hostile or attempting to cause harm.
    - **Measures of Self Perception and Sociability:**
        - **Self Depreciation:** Negative self-evaluation, extreme self-criticism, and feelings of inadequacy.
        - **Social Introversion:** Preference for solitude, minimal social participation, and discomfort in group settings.
    - **Measure of Critical Deviant Behaviors:**
        - **Deviation:** Endorsement of unusual, bizarre, or socially aberrant behaviors and unconventional experiences.
- **Scores and Scoring:**
    - **Raw Scores:** Calculated by summing endorsed items per scale based on scoring keys.
    - **Standard Scores/T-Scores:** Raw scores are converted into normative T-scores (typically anchored to standardized clinical and normal populations) to evaluate the severity of reported symptoms relative to peers.
    - **Profile Interpretation:** Displayed via visual profile sheets that highlight elevated scores indicating potential areas of clinical concern or vulnerability.
- **Administration Time:** typically requires approximately 30 to 35 minutes to complete the 240 true/false items.
- **Validity Scales and Response Bias**
    - **Denial Scale (Test-Taking Style):** Specifically designed to identify individuals who are overly guarded, defensive, or attempting to minimize/deny common human faults and psychological symptoms (detecting "fake-good" impression management).
    - **Inconsistency and Rare Item Checks:** Built-in safeguards to evaluate random responding, carelessness, or intentional malingering ("fake-bad").

Myers-Briggs Type Indicator (MBTI)

- **Purpose:** to make Jung's theory of psychological types understandable and practically useful in everyday life.
- **Population:** 14 years and older
- **Time:** requires approximately 15 to 30 minutes to complete the standard questionnaire.
- **Scales and Domains Measured:**
    - **Extraversion (E) or Introversion (I) –** ways to direct and receive energy
        - **Extroversion (E) –** gets energy from the outer world of people and experiences; focuses energy and attention outwards in action
        - **Introversion (I) –** gets energy from the inner world of reflections and thoughts; focuses energy and attention inwards in reflection
    - **Sensing (S) or Intuition (N) –** ways to take in information
        - **Sensing (S) –** prefers real information coming from the five senses; focuses on what is real
        - **Intuition (N) –** prefers information coming from associations; focuses on possibilities and what might be
    - **Thinking (T) or Feeling (F) –** ways to decide and come to conclusions
        - **Thinking (T) –** steps out of situations to analyze them dispassionately; prefers to make decisions on the basis of objective logic
        - **Feeling (F) –** steps into situations to weigh human values and motives; prefers to make decisions on the basis of values
    - **Judging (J) or Perceiving (P) –** ways to approach the outside world
        - **Judging (J) –** prefers to live life in a planned and organized manner; enjoys coming to closure and making a decision
        - **Perceiving (P) –** prefers to live life in a spontaneous and adaptable way; enjoys keeping options open

Edwards Personal Preference Schedule (EPPS)

- **Purpose:** to evaluate the relative importance of distinct motivational needs to help understand what drives an individual's behaviors, choices, and interpersonal style.
- **Population:** for adolescents and adults aged 16 to

REVIEWER NOTES BY ANNE
THIS IS NOT FOR SALE.
ANY FORMS OF REPRODUCTION
IS STRICTLY PROHIBITED.
rpm_anne | eireandtheclouds@gmail.com | x: rpm_anne | tiktok: eireandtheclouds

85+ years.

- **Scales and Domains Measured:**
  - **Achievement (ach):** To do one's best, to be successful, to accomplish tasks requiring skill and effort.
  - **Deference (def):** To get suggestions from others, to find out what others think, to follow instructions and conform to custom.
  - **Order (ord):** To write down nice clear lists, to keep things neat and organized, to plan ahead.
  - **Exhibition (exh):** To say witty and clever things, to talk about personal adventures, to be the center of attention.
  - **Autonomy (aut):** To be able to come and go as desired, to be independent of official constraints, to do things without regard to what others think.
  - **Affiliation (aff):** To be loyal to friends, to participate in friendly groups, to form strong attachments, to make as many friends as possible.
  - **Intraception (int):** To analyze one's motives and feelings, to observe others, to understand how feelings are caused.
  - **Succorance (suc):** To have others provide help when in trouble, to seek encouragement from others, to receive a lot of affection.
  - **Dominance (dom):** To be a leader in groups, to argue for one's point of view, to be regarded by others as a leader.
  - **Abasement (aba):** To feel guilty when one has done something wrong, to accept blame, to confess, to feel timid in the presence of superiors.
  - **Nurturance (nur):** To help friends, to assist others less fortunate, to treat others with sympathy and affection.
  - **Change (chg):** To do new and different things, to travel, to meet new people, to experience novelty and change in daily routine.
  - **Endurance (end):** To keep at a job until it is completely finished, to work hard at a task, to avoid being interrupted.
  - **Heterosexuality (het):** To go out with members of the opposite sex, to engage in social activities involving the opposite sex, to form romantic attachments.
  - **Aggression (agg):** To attack points of view that are contrary to one's own, to criticize others publicly, to tell others what one thinks of them.
- **Scores and Scoring**
  - **Raw Scores:** Calculated per need scale by summing the number of times statements representing that specific need are endorsed across the forced-choice pairs.
  - **Normative Percentiles:** Raw scores are mapped onto standardized percentile tables based on normative samples (typically categorized by gender, such as college student norms), displaying the relative strength of each need profile.
  - **Consistency Score:** Derived from repeated item pairs embedded in the test to check whether the examinee answered reliably across identical choices.
- **Administration Time:** requires approximately 35 to 45 minutes to complete the 225 forced-choice pairs.
- **Validity Scales and :**
  - **Forced-Choice Desirability Control:** Items within each pair are pre-tested to ensure equivalent social desirability ratings, minimizing the tendency for examinees to "fake good."
  - **Consistency Scale (Reliability Check):** Built-in duplicate item pairs (15 repeated pairs) are used to compute a Consistency Score, allowing the examiner to evaluate whether the respondent answered carelessly, randomly, or inconsistently throughout the inventory.

Sixteen Personality Factor Questionnaire (16PF)

- **Purpose:** to measure fundamental dimensions of normal personality to understand behavioral patterns across personal, educational, and occupational domains
- **Population:**
  - **Age Range:** Designed primarily for adolescents and adults aged 16 years and older.
  - **Reading Level:** Accessible at approximately a 5th- to 6th-grade reading level, ensuring broad applicability across varied educational backgrounds.
  - **Populations:** Widely utilized in corporate human resources, organizational development, executive coaching, career counseling, and psychological research.
- **Scales and Domains Measured**
  - **16 Primary Personality Scales:**
    - **Warmth (Factor A):** Outgoing, supportive, and attentive to others vs. reserved, detached,

STUDY NOTES BY ANNE
THIS IS NOT FOR SALE.
ANY FORMS OF REPRODUCTION IS STRICTLY PROHIBITED
fb: Anne Theclouds | x: rpm_anne | tiktok: eireandtheclouds

and cool.

- **Reasoning (Factor B):** Abstract-thinking, intelligent vs. concrete-thinking.
- **Emotional Stability (Factor C):** Calm, emotionally resilient vs. reactive, easily upset.
- **Dominance (Factor E):** Assertive, competitive, and forceful vs. cooperative, humble, and accommodating.
- **Liveliness (Factor F):** Enthusiastic, spontaneous vs. serious, restrained, and cautious.
- **Rule-Consciousness (Factor G):** Conscientious, dutiful, and conforming to rules vs. expedient, non-conforming.
- **Social Boldness (Factor H):** Uninhibited, socially bold, and adventurous vs. shy, threat-sensitive, and timid.
- **Sensitivity (Factor I):** Tender-minded, intuitive, and empathetic vs. utilitarian, objective, and tough-minded.
- **Vigilance (Factor L):** Suspicious, skeptical, and guarded vs. trusting, accepting, and unwary.
- **Abstractedness (Factor M):** Imaginative, idea-oriented, and absent-minded vs. practical, down-to-earth.
- **Privateness (Factor N):** Discreet, polished, and private vs. forthright, genuine, and artless.
- **Apprehension (Factor O):** Self-doubting, worried, and guilt-prone vs. self-assured, untroubled, and confident.
- **Openness to Change (Factor Q1):** Experimental, flexible, and analytical vs. traditional, attached to the familiar.
- **Self-Reliance (Factor Q2):** Self-sufficient, resourceful, and autonomous vs. group-oriented, dependent.
- **Perfectionism (Factor Q3):** Perfectionistic, organized, and self-disciplined vs. flexible, tolerating disorder.
- **Tension (Factor Q4):** Tense, driven, and impatient vs. relaxed, placid, and unfrustrated.
    - **Five Global Factors (Broad Domains):**
        - **Extraversion:** Combining warmth, liveliness, social boldness, privateness, and self-reliance.
        - **Anxiety:** Combining emotional stability, vigilance, apprehension, and tension.
        - **Tough-Mindedness:** Combining warmth, sensitivity, abstractedness, and openness to change.
        - **Independence:** Combining dominance, social boldness, vigilance, and openness to change.
        - **Self-Control:** Combining rule-consciousness, liveliness, abstractedness, and perfectionism.

- **Scores and Scoring**
    - **Raw Scores:** Calculated per scale based on scoring keys matching trait orientations.
    - **STEN Scores (Standard Ten):** Raw scores are converted into standard ten-point scales (with a Mean of 5.5 and a Standard Deviation of 2), allowing precise profile plotting across both primary traits and global factors.
    - **Normative Percentiles:** Scores are mapped against standardized demographic norms (stratified by gender, age, or occupational groups) to evaluate relative standing.
- **Administration Time:** requires approximately 35 to 50 minutes to complete the 185 items.
- **Validity Scales**
    - **Impression Management (IM):** Measures the tendency of an examinee to present themselves in an overly idealized, socially desirable, or "fake-good" light.
    - **Acquiescence (AC):** Evaluates the tendency to agree with items regardless of content, helping to catch indiscriminate or uncritical endorsement styles.
    - **Infrequency (INF):** Detects random responding, carelessness, or confusion by tracking unusual item choices rarely endorsed by normative samples.

Global Factors	Primary Factors

eireandtheclouds
REVIEWERS BY ANNE
THIS IS NOT FOR SALE.
ANY FORMS OF REPRODUCTION
IS STRICTLY PROHIBITED.
x: rpm_anne | tiktok: eireandtheclouds

Extraversion/ Introversion	(A) Warm-Reserved, (F) Lively-Serious, (H) Bold-Shy, (N) Private-Forthright, (Q2) Self-Reliant–Group-oriented
High Anxiety/ Low Anxiety	(C) Emotionally Stable–Reactive, (L) Vigilant–Trusting, (O) Apprehensive–Self-assured, (Q4) Tense–Relaxed
Tough-Mindedness/ Receptivity	(A) Warm–Reserved, (I) Sensitive–Unsentimental, (M) Abstracted–Practical, (Q1) Open-to-Change/Traditional
Independence/ Accommodation	(E) Dominant–Deferential, (H) Bold–Shy, (L) Vigilant–Trusting, (Q1) Open-to Change/ Traditional
Self-Control/ Lack of Restraint	(F) Lively–Serious, (G) Rule-conscious/Expedient, (M) Abstracted–Practical, (Q3) Perfectionistic–Tolerates disorder

NEO Five Factor Inventory (NEO-FFI)/NEO-PI-R/ NEO-PI-3

- **Purpose:** to provide a comprehensive, detailed, and scientifically rigorous assessment of an individual's normal-range personality traits across the lifespan
- **Population:** adults aged 18 years and older, with specialized normative tables and simplified item phrasing (particularly in the NEO-PI-3) enabling valid assessment of adolescents aged 12 to adult.
    - **Reading Level:** Accessible at approximately an 6th- to 8th-grade reading level, ensuring broad applicability across varied educational backgrounds.
    - **Populations:** Extensively utilized in clinical psychology, psychiatric practice, organizational development, executive coaching, and academic personality research globally.
- **Scales and Domains Measured:**
    - **Neuroticism (N)**: Adjustment vs. Emotional Instability
        - N1: Anxiety
        - N2: Angry Hostility
        - N3: Depression
        - N4: Self-Consciousness
        - N5: Impulsiveness
        - N6: Vulnerability
    - **Extraversion (E):** Interpersonal engagement and energy level
        - E1: Warmth
        - E2: Gregariousness
        - E3: Assertiveness
        - E4: Activity
        - E5: Excitement-Seeking
        - E6: Positive Emotions
    - **Openness to Experience (O):** Receptivity to novelty and intellectual curiosity
        - O1: Fantasy
        - O2: Aesthetics
        - O3: Feelings
        - O4: Actions
        - O5: Ideas
        - O6: Values
    - **Agreeableness (A):** Interpersonal orientation and prosocial behavior
        - A1: Trust
        - A2: Forthrightness
        - A3: Altruism
        - A4: Compliance
        - A5: Modesty
        - A6: Tender-Mindedness
    - **Conscientiousness (Con):** Impulse control, organization, and goal-directed persistence
        - Con1: Competence
        - Con2: Order
        - Con3: Dutifulness
        - Con4: Achievement Striving
        - Con5: Self-Discipline
        - Con6: Deliberation
- **Scores and Scoring**
    - **Raw Scores:** Calculated by summing item responses per domain and facet (adjusted for reverse-scored items).
    - **T-Scores:** Raw scores are converted into standardized T-scores (with a Mean of 50 and a Standard Deviation of 10), adjusted for age and gender norms.
    - **Profile Interpretation:** Displayed via visual profile sheets that categorize scores into descriptive bands (e.g., Very Low, Low, Average, High, Very High) relative to normative reference groups.
- **Administration Time**
    - **NEO-PI-3/NEO-PI-R (Full 240-item battery):** Typically requires approximately 30 to 40 minutes (up to 45 minutes) to complete.
    - **NEO-FFI (Short 60-item form):** Highly compressed, requiring approximately 10 to 15 minutes.
- **Validity Scales:**

eireandtheclouds STUDY NOTES BY ANNE THIS IS NOT FOR SALE. ANY FORMS OF REPRODUCTION IS STRICTLY PROHIBITED

- **(PPM) Positive Presentation Management –** intended to identify respondents claiming uncommon virtues and/or denying common faults
- **(NPM) Negative Presentation management –** intended to identify respondent claiming uncommon faults and/or denying common virtues
- **(INC) Inconsistency –** designed to detect random responding

Pictorial Self-Concept Scale for Children by Pasao

- **Purpose:** to evaluate a young child's self-perception, emotional adjustment, and self-esteem across multiple developmental contexts.
- **Population:** for young children in early childhood through lower elementary grades (typically ages 3 to 10 years/ kindergarten to grade 4).
- **Scales and Domains Measured:**
  - **Cognitive Competence:** Self-evaluation regarding learning ability, school performance, and intellectual capability.
  - **Physical Competence:** Perceptions of physical abilities, athletic skills, body coordination, and physical appearance.
  - **Peer Acceptance:** Feelings of belonging, popularity, and positive relationships with classmates and friends.
  - **Parental/Authority Acceptance:** Perceptions of affection, approval, and acceptance by parents and significant authority figures.
- **Scores and Scoring:**
  - **Raw Scores:** Calculated by assigning weighted values to the card placements (e.g., awarding points for positive self-image endorsements versus negative choices).
  - **Domain Subscores:** Aggregated scores reflecting specific areas such as cognitive competence, physical ability, and social interactions.
  - **Normative Standards:** Raw and composite scores are evaluated against age-appropriate normative tables to determine whether a child's self-concept falls within typical, high, or low developmental bands.
- **Administration Time:** requires approximately 10 to 20 minutes of individual administration in a quiet, engaging environment.
- **Validity Scales:**
  - **Consistency and Engagement Checks:** Examiners monitor the child's sorting consistency, attentiveness, and comprehension during the picture-sorting task to detect random selection, fatigue, or acquiescence bias.
  - **Cross-Informant Validation:** Comparing the child's pictorial self-concept profile with teacher or parent ratings to identify discrepancies between self-perception and observed behavior.

Panukat ng Ugali at Pagkatao (PUP by Enriquez)

- **Purpose:** designed to measure Filipino-oriented traits, behaviors, and attitudes, primarily to identify incentive talent
- **Population:** primarily for Filipino adolescents and adults across educational, corporate, and community settings.
- **Scales and Domains Measured**
  - **24 Trait Scales (141 items total):**
    - Ambisyon (Ambition)
    - Pagkamagalang (Respectfulness)
    - Katipiran (Thriftiness)
    - Katiyagaan (Perseverance)
    - Pagkamausisa (Inquisitiveness)
    - Hirap Kausapin/Pagkamayabang (Coyness or Unapproachability)
    - Lakas ng Loob (Guts / Daring)
    - Pagkamaalalahanin (Thoughtfulness)
    - Pagkamahiyain (Shyness/Timidity)
    - Pagkamalikhain (Creativity)
    - Pagkamapagbigay (Generosity)
    - Pagpapakumbaba (Humility)
    - Pagkamapagtimpi (Self-Control/Restraint)
    - Pagkamapunahin (Criticalness)
    - Pagkamaramdamin (Sensitiveness)
    - Pagkamatulong (Helpfulness)
    - Pagkapalaaway (Belligerence)
    - Pagkapikon (Low Tolerance for Teasing)
    - Pagkaresponsable (Responsibleness)
    - Pagkasalawahan (Fickle-Mindedness)
    - Pagkasigurista (Non-Risk-Taking/Prudence)
    - Pagkasunud-sunuran (Excessive Conformity)
    - Sumpong/Pagkama-mood

STUDY NOTES BY ANNE
THIS IS NOT FOR SALE.
ANY FORMS OF REPRODUCTION IS STRICTLY PROHIBITED.

(Moodiness)
    - Tigas ng Ulo (Stubbornness)
  - **Identifier/Criterion Items (19 items):** Items designed to capture specific culturally relevant behaviors and attitudes (such as gambling proneness, drinking habits, and prayer consistency) used primarily for research and validity tracking.
- **Administration Time:** requires approximately 30 to 45 minutes to complete.
- **Validity Scales:**
  - **Pagkakaila (Denial Scale):** Evaluates the respondent's tendency to deny basic human faults, common weaknesses, or universal truths (e.g., items concerning family obligations or personal shortcomings). Excessive endorsement invalidates the protocol.
  - **Kaugalian (Cultural Norms Scale):** Taps agreement with widely recognized traditional Philippine cultural values and behavioral norms to check whether the test taker is answering candidly or attempting to fake good/reject cultural realities.

Panukat ng Pagkataong Pilipino (PPP by Carlota)

- **Purpose:** comprehensively measure the important personality trait of a Filipino culture; it holds the idea of how people assess their personality characteristics, and understand which roles or appearance that greatly suits them
  - **Screens for the following:**
    - Research
    - Acceptance into program
    - Counseling
    - Employment
- **Population:** for adolescents and adults aged 13 years and older.
- **Scales and Domains Measured:**
  - **Form K (Porma K): Interpersonal Relationships (90 items across 8 subscales)**
    - Pagkamaalalahanin (Thoughtfulness)
    - Pagkamadaldal (Social Curiosity/Talkativeness)
    - Pagkamagalang (Respectfulness)
    - Pagkamaramdamin (Sensitiveness)
    - Pagkamasunurin (Obedience)
    - Pagkamatulungin (Helpfulness)
    - Pagkamaunawain (Capacity to be Understanding)
    - Pagkapalakaibigan (Sociability/Friendliness)
  - **Form S (Porma S): Personal Traits (80 items across 7 subscales)**
    - Pagkamaayos/Orderliness
    - Pagkamahinahon/Emotional Stability
    - Pagkamapagkumbaba/ Humility
    - Pagkamasayahin/ Cheerfulness
    - Pagkamatapat/Honesty
    - Pagkamatiyaga/Patience
    - Pagkaresponsable/ Responsibility
  - **Form KS (Porma KS): Intelligence and Creativity (40 items across 4 subscales)**
    - Pagkamalikhain/Creativity
    - Pagkamapagsapalaran/Risk Taking
    - Pagkamasikap/Achievement Orientation
    - Pagkamatalino/Intelligence
- **Scores and Scoring:**
  - **Raw Scores:** Calculated by summing item responses across each subtest scale.
  - **Normative Standards:** Raw totals are converted into Percentile Ranks and Normalized Standardized Scores based on normative reference tables.
- **Administration Time:**
  - **Form K or Form S:** Approximately 20 to 30 minutes each.
  - **Form KS:** Approximately 10 to 20 minutes.
  - **Full Battery (Forms K, S, and KS combined):** Approximately 50 to 60 minutes.

C | Interest Tests

Strong Campbell Interest Inventory Test

- **Purpose:** to guide individuals in career exploration, educational planning, and vocational counseling
- **Population:** adolescents (high school age, approx. 14+ years), college students, and adults exploring career paths or transitions.
  - **Reading Level:** Accessible at approximately an 8th-grade reading level, ensuring broad applicability across educational backgrounds.
  - **Populations:** Widely utilized in university career centers, high school guidance departments, corporate outplacement services, executive coaching, and vocational rehabilitation.

eireandtheclouds NOTES BY ANNE
THIS IS NOT FOR SALE.
ANY FORMS OF REPRODUCTION IS STRICTLY PROHIBITED.

- **Scales and Domains Measured:**
  - **General Occupational Themes (GOTs): The broad, foundational RIASEC dimensions reflecting an individual's core orientation toward work and life:**
    - **Realistic (R):** Hands-on, practical, mechanical, and physical activities.
    - **Investigative (I):** Analytical, scientific, intellectual, and problem-solving pursuits.
    - **Artistic (A):** Creative, expressive, imaginative, and aesthetic endeavors.
    - **Social (S):** Helping, teaching, supporting, and counseling others.
    - **Enterprising (E):** Leading, persuading, managing, and selling.
    - **Conventional (C):** Organizing, structuring, managing data, and adhering to systematic procedures.
  - **Basic Interest Scales (BISs):** Finer-grained interest clusters nested within each GOT (measuring specific areas such as Public Speaking, Mathematics, Nature/Agriculture, or Art), pinpointing exact areas of engagement.
  - **Occupational Scales (OSs):** Hundreds of specific job titles comparing the respondent's similarity of interests to satisfied workers currently employed in those exact professions.
  - **Personal Style Scales (PSSs):** Measures broad lifestyle and working orientations:
    - **Work Style:** Preference for working with people versus data/things.
    - **Learning Environment:** Comfort with academic versus practical learning settings.
    - **Leadership Style:** Tendency to lead through influence, direction, or collaboration.
    - **Risk Taking:** Willingness to embrace physical, financial, or intellectual risks.
    - **Team Orientation:** Preference for autonomous work versus collaborative team structures.
- **Scores and Scoring**
  - **Raw Scores:** Calculated per scale by summing weighted responses matching specific interest keys.
  - **Standard Scores (T-Scores):** Raw scores are converted into standardized T-scores (with a Mean of 50 and a Standard Deviation of 10), normed separately by gender to account for historical interest differences while providing a clear relative standing.
  - **Profile Interpretation:** Displayed via visual profile sheets showing high, average, or low scores across the 6 GOTs, dozens of BISs, and specific Occupational Scales, indicating the degree of similarity or alignment.
- **Administration Time:** requires approximately 20 to 30 minutes to complete the online or paper questionnaire.
- **Versions and Forms**
  - Strong Vocational Interest Blank (SVIB)
  - Strong-Campbell Interest Inventory (SCII)
  - Modern Strong Interest Inventory (SII)

Thurstone Interest Schedule

- **Purpose:** to identify an individual's preferred occupational directions and assist in career exploration
- **Population:** or adolescents (secondary and high school students), college undergraduates, and adults navigating career transitions or vocational exploration.
- **Scales and Domains Measured:**
  - **Outdoor:** Interest in work involving nature, plants, animals, and physical outdoor activities.
  - **Mechanical:** Preference for working with tools, machines, equipment, and structural systems.
  - **Computational:** Enjoyment of numerical operations, mathematics, and quantitative data management.
  - **Scientific:** Orientation toward scientific investigation, laboratory experimentation, and analytical problem-solving.
  - **Persuasive:** Attraction to leadership, sales, public speaking, and influencing others.
  - **Artistic:** Creativity and appreciation for design, visual arts, and aesthetic expression.
  - **Literary:** Interest in reading, writing, language structure, and communication arts.
  - **Musical:** Interest in music appreciation, composition, performance, and auditory

STUDY NOTES BY ANNE | NOT FOR SALE | ANY FORMS OF REPRODUCTION IS STRICTLY PROHIBITED

arts.
    - **Social Service:** Desire to help, guide, teach, and support others in community or personal contexts.
    - **Clerical:** Preference for structured office routines, organization, filing, and systematic administrative tasks.
- **Administration Time:** requires approximately 15 to 20 minutes to complete the 100 paired items.

OTHER PSYCHOLOGICAL TESTS THAT ARE ALSO COMMONLY* USED IN THE PHILIPPINES (not mentioned in Annex B)

A. Achievement Tests

Wide Range Achievement Test (WRAT-5) – is an individually administered achievement test designed to assess fundamental academic skills in reading, spelling, and math computation

- **Screening and Diagnosis:** To identify learning disabilities, academic delays, or discrepancies between intellectual potential and actual academic achievement (ability-achievement discrepancy models).
- **Progress Monitoring:** To track academic growth or deterioration over time in clinical or educational interventions.
- **Re-evaluations:** To provide baseline data and follow-up measures for individuals with neurological injuries, learning disorders, or cognitive changes.
- **Population:** 5 to 85+ years.
- **Time:**
    - **Younger children (ages 5–7):** Typically 15 to 25 minutes.
    - **Older children, adolescents, and adults:** Typically 30 to 45 minutes.
- **Scores and Scoring:**
    - **Standard Scores:** Provided with a mean of 100 and a standard deviation of 15, allowing comparisons across age groups.
    - **Percentile Ranks:** Indicate the percentage of individuals in the normative sample who scored at or below a given score.
    - **Stanines and Grade/Age Equivalents:** Offer additional developmental benchmarks to interpret performance relative to peers.
    - **Composite Scores:**
        - **Math-Reading Composite:** Combines math computation and reading skills.
        - **Reading Composite:** Combines word reading and sentence comprehension (where applicable).
        - **Total Achievement Score:** An overall index summarizing performance across all core subtests.
- **Scales and Domains Measured (Subtests)**
    - **Word Reading:** Measures letter identification and word-recognition skills through pronunciation of printed words out of context.
    - **Sentence Comprehension:** Measures the ability to read shallow-to-complex sentences and understand meaning by supplying missing words (cloze procedure).

      Note: This is a newer addition introduced to better capture reading comprehension rather than just decoding.
    - **Spelling:** Measures an individual's ability to write letters and words from dictation.
    - **Math Computation:** Measures mathematical skills ranging from counting and simple addition to advanced operations involving fractions, algebra, and geometry.

B. Intelligence Tests

SRA Verbal and Nonverbal Tests – are standardized group or individually administered cognitive assessment batteries.

- **Verbal Batteries – t**ypically assess word fluency, verbal meaning, and reasoning through language.
- **Nonverbal Batteries –** evaluate spatial visualization, perceptual speed, and abstract reasoning independent of extensive reading or language proficiency.
- **Purpose:**
    - **Educational Placement and Guidance:** To assist schools in academic streaming, gifted program identification, and vocational tracking.
    - **Cognitive Profiling:** To identify specific strengths and weaknesses across different mental abilities (e.g., distinguishing between verbal comprehension and spatial manipulation).
    - **Screening:** To serve as a quick, reliable predictor of academic success and general learning potential without requiring a lengthy individual intelligence test (like the WAIS or WISC).
- **Population:** through adolescence (high school) and into adult vocational assessment.
- **Time:** typically ranges from 30 to 60 minutes depending on the specific sub-tests administered

eireandtheclouds STUDY NOTES BY ANNE THIS IS NOT FOR SALE ANY FORMS OF REPRODUCTION IS STRICTLY PROHIBITED

and the age level.

- Scores and Scoring:
  - **Raw Scores:** Total number of correct answers within the allotted time limits.
  - **Standard Scores/IQ Equivalents:** Converted to normalized standard scores (often with varying means depending on the specific historical SRA edition, though commonly mapped to standard metrics like deviation IQs or percentiles).
  - **Percentile Ranks:** Used to compare an individual's performance against national normative samples stratified by age or grade.
  - **Mental Ages:** Older editions frequently utilized mental age equivalents, though modern interpretations rely heavily on standard scores and stanines.
- **Scales and Domains Measured:**
  - **Verbal Meaning (V):** Comprehension of ideas expressed through words, vocabulary knowledge, and verbal reasoning.
  - **Word Fluency (W):** Ability to write or recall words rapidly based on lexical constraints (e.g., words starting with a specific letter).
  - **Space/Spatial Visualization (S):** Capacity to mentally manipulate geometric shapes, visualize objects in two or three dimensions, and understand spatial orientations (a core nonverbal domain).
  - **Reasoning (R)/Logical Ability:** Ability to solve logical problems, foresee consequences, and identify patterns or rules in abstract series (nonverbal reasoning).
  - **Number/Numerical Facility (N):** Speed and accuracy in handling simple arithmetic computations.
  - **Perceptual Speed (P):** Ability to recognize visual similarities and differences quickly.

Test of Nonverbal Intelligence, Fourth Edition (TONI-4) – is an individually administered intelligence test that assesses fluid intelligence and general cognitive capacity (g) using abstract visual stimuli.

- **Purpose:**
  - **Fair Assessment for Impaired Populations:** To evaluate cognitive ability in individuals with language, hearing, or motor impairments, autism, traumatic brain injury, stroke, or intellectual disabilities.
  - **Non-English/Culturally Diverse Testing:** To assess intelligence in individuals who are not proficient in English or who come from diverse cultural and linguistic backgrounds.
  - **Cognitive Screening:** To provide a reliable estimate of general problem-solving ability and abstract reasoning without the confounding variable of verbal proficiency.
- **Population:** 66 years, 0 months to 89 years, 11 months.
- **Time:** 15 to 20 minutes to complete.
- **Scores and Scoring:**
  - **Index Scores (TONI-4 Quotient):** Standard scores typically scaled with a mean of 100 and a standard deviation of 15.
  - **Percentile Ranks:** Indicate relative standing compared to normative peers.
  - **Age Equivalents:** Provide developmental growth markers based on mental age comparisons.
- **Domains Measured:** abstract problem-solving and fluid reasoning through items incorporating one or more of the following visual characteristics:
  - Shape
  - Position
  - Direction
  - Rotation
  - Contiguity
  - Shading
  - Size
  - Movement

Kaufman Assessment Battery for Children, Second Edition Normative Update (KABC-II NU) – is a comprehensive psychometric assessment designed to evaluate how children receive and process information

- **Purpose:**
  - **Culturally Fair Assessment:** To evaluate cognitive capacity in children from diverse cultural, ethnic, and linguistic backgrounds (including bilingual or English as a Second Language learners) with minimal bias.
  - **Cognitive Profiling:** To identify specific patterns of cognitive strengths and weaknesses to help explain learning difficulties and guide targeted educational interventions.
  - **Clinical and Educational Planning:** To provide actionable insights into information processing styles, assisting in special education placement, neuropsychological evaluations, and treatment planning.
- **Population:** 3 years, 0 months to 18 years, 11 months

Study Notes by Anne | NOT FOR SALE. ANY FORMS OF REPRODUCTION IS STRICTLY PROHIBITED

- Time:
  - **Luria Model Core Battery:** Typically takes 25 to 60 minutes.
  - **CHC Model Core Battery:** Typically takes 30 to 75 minutes (depending on the examinee's age).
- **Scores and Scoring:**
  - **Global Composite Scores:**
    - Mental Processing Index (MPI) for the Luria model.
    - Fluid-Crystallized Index (FCI) for the CHC model (note: the FCI can be substituted with a Nonverbal Index when verbal or motor constraints require it).
  - **Scale and Subtest Scaled Scores:** Scaled scores with a mean of 10 and standard deviation of 3.
  - **Standard Scores:** Index scores scaled with a mean of 100 and standard deviation of 15.
  - **Percentile Ranks and Age Equivalents:** Provide normative comparisons and developmental markers.
- **Scales and Domains Measured:**
  - **Luria's Neuropsychological Model (Processing Scales)**
    - **Sequential Processing (Gsm):** Solving problems by serial or step-by-step organization (Short-Term Memory).
    - **Simultaneous Processing (Gv):** Solving problems by integrating stimuli all at once (Visual Processing).
    - **Planning (Gf):** Solving novel problems using executive function, reasoning, and strategic planning (Fluid Reasoning).
    - **Learning (Glr):** Storing and efficiently retrieving new information (Long-Term Storage and Retrieval).
  - **Cattell-Horn-Carroll (CHC) Psychometric Model**
    - Extends the Luria scales by incorporating Crystallized Ability (Gc) (Verbal Knowledge), offering a comprehensive broad-ability cognitive profile.
- **Subtests:**
  - **Triangles/Block Counting:** Visual-spatial manipulation and construction.
  - **Face Recognition/Gestalt Closure:** Perceptual organization and closure.
  - **Number Recall/Word Order:** Sequential memory and auditory processing.
  - **Atlantis/Atlantis Delayed:** Associative learning and long-term retrieval.
  - **Pattern Reasoning/Conceptual Thinking:** Fluid reasoning and categorization.
  - **Verbal Knowledge/Expressive Vocabulary:** Language development and crystallized knowledge (Gc, restricted to the CHC model).
  - **Hand Movements/Rover:** Motor planning and spatial execution.

Woodcock-Johnson V (WJ V) – it evaluates an individual's cognitive profile, academic capabilities, and oral language skills across an exceptionally wide age span.

- **Purpose:**
  - **Comprehensive Psychoeducational Evaluation:** To provide deep diagnostic insights into how an individual processes information, learns academic concepts, and utilizes language.
  - **Identification of Specific Learning Disabilities (SLDs):** To evaluate ability-achievement discrepancies and assess cognitive processing weaknesses across all recognized SLD domains.
  - **Instructional and Clinical Planning:** To identify individual instructional zones and generate data-driven strategies for educational intervention, gifted programming, or clinical rehabilitation.
- **Population:** from early childhood (ages 3–4) through adulthood and geriatrics (up to 90+ years).
- **Time:**
  - **Core Cognitive Batteries:** Typically 30 to 60 minutes.
  - **Achievement Batteries:** Typically 45 to 75 minutes.
- **Scores and Scoring:**
  - **Standard Scores:** Scaled with a mean of 100 and a standard deviation of 15.
  - **Percentile Ranks:** Indicate relative standing within the normative sample.
  - **Age and Grade Equivalents:** Provide developmental growth markers.
  - Cluster Scores: Broad and narrow ability groupings that offer high reliability for clinical interpretation.
  - **Automated Scoring:** Real-time platform management of basal/ceiling rules, dynamic re-routing, and voice-capture review features.

eireandtheclouds | STUDY NOTES BY ANNE | NOT FOR SALE. ANY FORM OF REPRODUCTION IS STRICTLY PROHIBITED.

- **Scales and Domains Measured:**
  - **WJ V Tests of Cognitive Abilities (COG):** Measures broad and narrow cognitive functions, including:
    - Comprehension-Knowledge
    - Fluid Reasoning
    - Short-Term Working Memory
    - Long-Term Storage and Retrieval
    - Visual Processing
    - Auditory Processing
    - Processing Speed
  - **WJ V Tests of Achievement (ACH):** Measures foundational and advanced academic performance in reading, mathematics, written expression, and academic knowledge.
  - **Oral Language & Virtual Test Library (VTL):** Measures listening comprehension, oral expression, and specialized linguistic structures.
- **Subtests**
  - **Number Series/Concept Formation/Analysis-Synthesis:** Fluid reasoning and logical deduction.
  - **Verbal Attention/Numbers Reversed/Object-Number Sequencing:** Working memory and executive processing.
  - **Phonological Processing/Sound Blending:** Auditory processing and dyslexia risk factors.
  - **Word Reading/Sentence Comprehension/Math Problem Identification:** Core academic decoding, reading comprehension, and applied problem-solving.

C. Aptitude tests

Otis-Lennon School Ability Test (OLSAT-8) – measures cognitive capability, abstract thinking, and logical deduction, making it a primary instrument for educational tracking and gifted and talented program identification.

- **Purpose:**
  - **Gifted and Talented Identification:** To identify students with superior reasoning and cognitive capacities who qualify for advanced or accelerated academic programs.
  - **Ability vs. Achievement Comparison:** To compare a student's general intellectual potential against their actual classroom achievement scores (such as the Stanford Achievement Test series).
  - **Educational Guidance:** To assist educators in tailoring instructional approaches to match individual learning styles and cognitive strengths.
- **Population:** from Pre-Kindergarten through Grade 12, organized across seven grade-specific levels (Levels A through G).
- **Time:** takes 40 to 75 minutes depending on the test level and student age.
- **Scores and Scoring:**
  - **School Ability Index (SAI):** A normalized standard score with a mean of 100 and a standard deviation of 16, indicating relative standing against age peers.
  - **Percentile Ranks:** Show the percentage of students in the normative sample scoring at or below a given level.
  - **Stanines:** Provide broad categorical bands for group comparisons.
  - **Verbal and Nonverbal Scores:** Sub-composite scores that delineate verbal processing strengths from nonverbal and quantitative reasoning capacities.
- **Scales and Domains Measured:**
  - **Verbal Area**
    - **Verbal Comprehension:** Assessing listening comprehension, following directions, and understanding word/sentence relationships.
    - **Verbal Reasoning:** Evaluating verbal analogies, classifications, and logical selections using language structures.
  - **Nonverbal Area**
    - **Pictorial Reasoning:** Solving visual problems using pictures and diagrams (heavily weighted in early levels).
    - **Figural Reasoning:** Assessing spatial orientation, geometric transformations, and pattern recognition using shapes.
    - **Quantitative Reasoning:** Measuring mathematical reasoning, number series, and numeric matrix completions.
- **Subtests:**
  - **Following Directions/Aural Reasoning:** Listening-based problem solving (early levels).
  - **Picture/Figural Classification & Analogies:** Identifying similarities, differences, and transformations across

eireandtheclouds NOTES BY ANNE | NOT FOR SALE. ANY FORMS OF REPRODUCTION IS STRICTLY PROHIBITED | x: rpm_anne | tiktok: eireandtheclouds

visual stimuli.
    - **Number/Word Matrices:** Deducing logical missing pieces in a grid based on intersecting rules.
    - **Arithmetic Reasoning & Number Series:** Applying computational and structural logic to sequence problems.
- **Versions:**
    - Kindergarten: Level A
    - Grade 1: Level B
    - Grade 2: Level C
    - Grade 3: Level D
    - Grade 4 to 5: Level E
    - Grade 6 to 8: Level F
    - Grade 9 to 12: Level G

D. Normal Personality Tests

California Psychological Inventory (CPI-434) – it is grounded in "folk concepts"—natural, culturally universal descriptors that people use to evaluate the behavior of others in daily life.

- **Purpose:**
    - **Personal and Professional Development:** To provide insight into leadership potential, managerial style, interpersonal effectiveness, and interpersonal strengths.
    - **Pre-Employment Screening & Selection:** Widely used in industrial-organizational psychology for high-stakes screening (e.g., law enforcement, public safety, corporate executive positions) to assess candidate reliability, judgment, and emotional stability.
    - **Counseling and Psychotherapy:** To guide individual counseling by clarifying personal values, social interaction patterns, and life goals.
- **Population:** typically starting around age 13 through adulthood
- **Time:** takes 45 to 60 minutes to complete.
- **Scores and Scoring:**
    - **Standard Scores (T-Scores):** Standardized with a mean of 50 and a standard deviation of 10, allowing clear comparisons against relevant normative groups (e.g., general population or professional occupational groups).
    - **Vector Scales:** Three major structural vectors (V1: Internal vs. External orientation, V2: Rule-favoring vs. Rule-questioning, V3: Self-realization/fulfillment) that map out an individual's broad psychological orientation.
    - **Profile Analysis:** Interpreted through standardized profile sheets that plot scores across folk concepts to generate configural patterns rather than relying on single isolated scores.
- **Scales and Domains Measured:**
    - **Class I: Measures of Poise, Self-Assurance, Interpersonal Adequacy, and Innovativeness**
        - Dominance (Do)
        - Capacity for Status (Cs)
        - Sociability (Sy)
        - Social Presence (Sp)
        - Self-Acceptance (Sa)
        - Independence (In)
        - Empathy (Em)
    - **Class II: Measures of Socialization, Maturity, Responsibility, and Structuring Behavior**
        - Responsibility (Re)
        - Socialization (So)
        - Self-Control (Sc)
        - Good Impression (Gi)
        - Communality (Cm)
        - Well-Being (Wb)
        - Tolerance (To)
    - **Class III: Measures of Achievement Potential and Intellectual Efficiency**
        - (Achievement via Conformance (Ac)
        - Achievement via Independence (Ai)
        - Intellectual Efficiency (Ie)
    - **Class IV: Measures of Intellectual and Conceptual Modes**
        - Psychological-Mindedness (Py)
        - Flexibility (Fx)
        - Femininity/Masculinity (F/M)

E. Psychopathology Personality Tests

Minnesota Multiphasic Personality Inventory-3 (MMPI-3) – is the most recent, empirically updated version of the premier, standardized psychometric test used to assess psychopathology, personality traits, and psychological functioning.

- **Purpose:**
    - **Clinical Diagnosis:** To assist mental health clinicians in identifying and diagnosing mental disorders, personality pathologies, and emotional disturbances in clinical settings.
    - **Forensic and Legal Evaluations:** Widely used in court-ordered evaluations, child custody disputes, personal injury cases, and risk assessments due to its robust

SELF-REVIEWER NOTES BY ANNE
THIS IS NOT FOR SALE.
ANY FORMS OF REPRODUCTION IS STRICTLY PROHIBITED.

validity scales that detect defensive responding, malingering, or random answers.
    - **Pre-Employment Screening:** Utilized in high-stakes public safety and occupational screening (e.g., law enforcement, aviation, nuclear power) to evaluate psychological fitness for duty.
    - **Treatment Planning:** To guide psychotherapy and clinical interventions by providing insight into an individual's internal distress, interpersonal patterns, and behavioral tendencies.
- **Population:** 18 years and older
- **Time:** 35 to 50 minutes
- **Scores and Scoring:**
    - **T-Scores:** Standardized scores with a mean of 50 and a standard deviation of 10, allowing clinicians to interpret elevations relative to modern normative reference groups (non-clinical normative samples).
    - **Restructured Clinical (RC) Scales & Subscales:** Core interpretive scales that measure distinct clinical constructs while minimizing overlap and intercorrelation.
    - **Profile Configurations:** Interpreted via computer-generated profile reports that map scores across validity, higher-order, restructured clinical, and specific problem scales.
- **Scales and Domains Measured:**
    - **Validity Scales**
        - **Inconsistency Scales:** Variable Response Inconsistency (VRIN-r), True Response Inconsistency (TRIN-r).
        - **Under-reporting Scales:** Unlikely Virtues (F-r, Fp-r, Fs), Adjustment Validity (L-r, K-r).
        - **Over-reporting Scales:** Infrequent Psychopathology Responses (Fp-r), Symptom Validity (FBS-r).
    - **Higher-Order (HOC) Scales**
        - **Emotional/Internalizing Dysfunction (EID):** Measures problems involving depression, anxiety, and general emotional distress.
        - **Thought Dysfunction (THD):** Measures psychotic-like thinking, unusual perceptions, and disorganized cognitive processes.
        - **Behavioral/Externalizing Dysfunction (BXD):** Measures substance abuse, antisocial behavior, and impulsivity.
    - **Restructured Clinical (RC) Scales**
        - **Demoralization (RCd):** General unhappiness and dissatisfaction.
        - **Somatic Complaints (RC1):** Diffuse physical symptoms.
        - **Low Positive Emotions (RC2):** Lack of zest, anhedonia, and depression.
        - **Cynicism (RC3):** Distrust of human motives.
        - **Antisocial Behavior (RC4):** Rule-breaking and aggressive impulses.
        - **Ideas of Persecution (RC6):** Self-referential beliefs and paranoia.
        - **Dysfunctional Negative Emotions (RC7):** Anxiety, irritability, and vulnerability.
        - **Aberrant Experiences (RC8):** Unusual sensory and cognitive experiences.
        - **Hypomanic Activation (RC9):** Elevated energy, impulsivity, and grandiosity.
    - **Specific Problems (SP) Scales**
        - Detailed subscales capturing specific psychological domains such as somatic/cognitive symptoms (malaise, neurological complaints), internalizing problems (suicidal ideation, helplessness, alienation, anxiety, fears), and externalizing problems (substance abuse, juvenile conduct problems, aggression).
    - **Versions:**
        - **MMPI-3:** The current modern standard edition featuring 335 items, updated contemporary normative samples, and modernized linguistic phrasing.
        - **MMPI-2/MMPI-2-RF:** Previous legacy editions containing 567 items (MMPI-2) or a restructured 338-item form (MMPI-2-RF). While historically widespread, the MMPI-3 represents the current clinical gold standard.
        - **MMPI-A-RF:** A specialized

STUDY NOTES BY ANNE | eireandtheclouds
THIS IS NOT FOR SALE. ANY FORM OF REPRODUCTION IS STRICTLY PROHIBITED.
eireandtheclouds | x: rpm_anne | tiktok: eireandtheclouds

241-item version tailored specifically for adolescent evaluation (ages 14–18).

Millon Clinical Multiaxial Inventory-IV (MCMI-IV) – is the latest iteration of the premier, theory-based self-report clinical inventory designed to evaluate personality psychopathology and clinical syndromes in adult psychiatric populations

- **Purpose:**
  - **Clinical Assessment and Diagnosis:** To assist mental health professionals in evaluating underlying personality pathology, emotional disorders, and severe clinical syndromes.
  - **Treatment Planning:** To provide deep insight into a patient's coping styles, interpersonal dynamics, and treatment vulnerabilities, thereby guiding therapeutic interventions.
  - **Forensic and Legal Evaluations:** To aid in court evaluations, pre-sentence investigations, and risk assessments where personality dynamics and psychopathology are central to the case.
- **Population:** 18 years and older who are currently undergoing psychological evaluation or psychiatric treatment.
- Time: 25 to 30 minutes to complete.
- **Score and Scoring:** unique scoring system called **Base Rate (BR) scores—**reflect the prevalence of a particular disorder in the clinical normative sample
  - A **BR score of 60** is the median, but scores below 60 are generally not interpreted
  - A **BR score of 75** is the cutoff for the presence of a disorder, indicating a significant elevation
  - A **BR score of 85** or higher suggests the disorder is **prominent** in the individual's profile
- **Scales and Domains Measured**
  - **Validity Scales**
    - **Modifying Indices (Disclosure, Desirability, Debasement):** Evaluates test-taking attitude, defensive minimization, frankness, or exaggeration/cry-for-help responding.
    - **Inconsistency and Random Response Indicators:** Ensures data integrity by flagging invalid or contradictory answering patterns.
  - **Clinical Personality Patterns**
    - **Measures enduring, pervasive structural personality styles:** Schizoid, Avoidant, Melancholic/Depressive, Dependent, Histrionic, Narcissistic, Antisocial, Aggressive/Sadistic, Compulsive, Negativistic (Passive-Aggressive), and Masochistic (Self-Defeating).
  - **Severe Personality Pathology**
    - **Measures profound personality disorganization:** Schizotypal, Borderline, and Paranoid personality scales.
  - **Clinical Syndromes**
    - **Measures acute emotional and psychological distress states:** Generalized Anxiety, Somatic Symptom, Bipolar Spectrum, Persistent Depression, Alcohol Use, Drug Use, Thought Disorder, Major Depression, and Delirium/Severe Cognitive Impairment.

Personality Assessment Inventory (PAI) – is a standardized, objective self-report questionnaire of personality and psychopathology designed to provide comprehensive information on clinical status and psychological functioning.

- **Purpose:**
  - **Clinical Assessment and Diagnosis:** To assist mental health professionals in evaluating psychopathology, personality features, and emotional distress.
  - **Treatment Planning:** To provide actionable data on therapeutic issues, symptom severity, interpersonal style, and potential treatment complications (such as aggression or risk of self-harm).
  - **Forensic and Correctional Evaluations:** Extensively used in legal contexts, court-ordered evaluations, pre-sentence investigations, and risk assessments due to its robust validity scales that detect malingering, defensiveness, or random responding.
- **Population:** 18 years and older (a specialized adolescent version, the PAI-A, exists for ages 12–18).
- **Time:** 40 to 50 minutes
- **Scores and Scoring**
  - **T-Scores:** Standardized scores based on a national normative reference group, scaled to have a mean of 50 and a

eireandtheclouds
STUDY NOTES
NOT FOR SALE.
ANY FORMS OF REPRODUCTION
IS STRICTLY PROHIBITED.
eireandtheclouds@gmail.com | x: rpm_anne | tiktok: eireandtheclouds

standard deviation of 10. T-scores of 70 or higher generally indicate clinically significant elevations.
    - **Reliable Change Index (RCI):** Facilitates tracking symptom change over time across pre- and post-treatment evaluations.
    - **Profile Reports:** Computer-generated narrative reports that plot scores across validity scales, clinical domains, and treatment considerations.
- **Scales and Domains Measured**
    - **Validity Scales (4 Scales)**
        - **Inconsistency (ICN):** Detects random or careless responding.
        - **Infrequency (INF):** Detects confusion, eccentric responding, or gross carelessness.
        - **Negative Impression (NIM):** Measures exaggeration, cry-for-help responding, or malingering.
        - **Positive Impression (PIM):** Measures defensiveness, minimization of flaws, or attempts to present in an overly favorable light.
    - **Clinical Scales (11 Scales)**
        - **Somatic Complaints (SOM):** Conversion, somatization, and health concerns.
        - **Anxiety (ANX):** Cognitive, affective, and physiological anxiety.
        - **Anxiety-Related Disorders (ARD):** Phobias, obsessive-compulsive traits, and traumatic stress.
        - **Depression (DEP):** Cognitive, affective, and physiological signs of depression.
        - **Manic (MAN):** Activity level, grandiosity, and irritability.
        - **Paranoia (PAR):** Hypervigilance, persecution, and resentment.
        - **Schizophrenia (SCZ):** Psychotic experiences, social detachment, and thought disorder.
        - **Borderline Features (BOR):** Affective instability, identity problems, negative relationships, and self-harm.
        - **Antisocial Features (ANT):** Antisocial behaviors, egocentricity, and stimulus-seeking.
        - **Alcohol Problems (ALC):** Consequences and habits related to alcohol use.
        - **Drug Problems (DRG):** Consequences and habits related to drug use.
    - **Treatment Consideration Scales (5 Scales)**
        - **Aggression (AGG):** Aggressive attitude, verbal aggression, and physical potential.
        - **Suicidal Ideation (SUI):** Thoughts and plans concerning self-harm.
        - **Stress (STR):** Recent environmental stressors and life crises.
        - **Nonsupport (NON):** Lack of perceived social support.
        - **Treatment Rejection (RXR):** Lack of interest in making psychological changes or resistance to treatment.
    - **Interpersonal Scales (2 Scales)**
        - **Dominance (DOM):** Independence and control in relationships.
        - **Warmth (WRM):** Empathy, interest in others, and interpersonal affiliation.

F. Attitude Test

Satisfaction with Life Scale

- **Purpose:** Unlike measures that assess positive and negative affect or specific life domains (such as health, finances, or work), the SWLS focuses on a person's conscious evaluative judgment of their life as a whole.
    - **Subjective Well-Being Assessment:** To evaluate an individual's overall cognitive satisfaction with life as a primary component of subjective well-being.
    - **Research and Clinical Screening:** Widely utilized in positive psychology, sociological studies, counseling, and therapeutic contexts to track life satisfaction outcomes, monitor intervention efficacy, and assess general psychological adjustment.
    - **Life Quality Evaluation:** To provide a quick, reliable indicator of perceived quality of life across varied research and applied settings.

eireandtheclouds
REVIEWER NOTES BY ANNE
THIS IS NOT FOR SALE.
ANY FORMS OF REPRODUCTION
IS STRICTLY PROHIBITED.
eireandtheclouds@gmail.com | x: rpm_anne | tiktok: eireandtheclouds

- **Population:** 10 years and older
- **Time:** Exceptionally brief; typically takes 1 minute (or less than 2 to 3 minutes) to complete.
- **Scores and Scoring:**
  - **Total Score Calculation:** The responses to the 5 items are summed together to yield a total score ranging from 5 to 35.
  - **Score Interpretation Benchmarks:**
    - 31–35: Extremely satisfied
    - 26–30: Satisfied
    - 21–25: Slightly satisfied
    - 20: Neutral
    - 15–19: Slightly dissatisfied
    - 10–14: Dissatisfied
    - 5–9: Extremely dissatisfied
  - **Normative Comparison:** Higher scores indicate a higher perceived level of life satisfaction, while lower scores reflect significant dissatisfaction and potential psychological distress.
- **Scales and Domains Measured:**
  - **Global Life Satisfaction:** The cognitive evaluation of one's life against personal standards, rather than emotional mood states or domain-specific satisfactions.
- **Subtests:**
  - The SWLS does not feature subtests or multifaceted subscales; it consists of a single unified scale comprising 5 core statements:
    - In most ways my life is close to my ideal.
    - The conditions of my life are excellent.
    - I am satisfied with my life.
    - So far I have gotten the important things I want in life.
    - If I could live my life over, I would change almost nothing.

G. Projective Tests

Rotter Incomplete Sentence Blank (RISB-2)

- **Purpose:** focus on assessing psychological adjustment high school, college, and adults
- **Population:** high school, college, and adults
- **Time:** 20 to 40 minutes
- **Domains Measured:** psychological maladjustment

Sachs Sentence Completion Test (SSCT)

- **Purpose:** it aims to uncover an individual's attitudes, feelings, and conflicts by having them complete incomplete sentences
- **Population:** adults
- **Time:** 20 to 40 minutes
- **Domains Measured:**
  - Family
    - Attitude towards Mother
    - Attitude towards Father
    - Attitude towards Family Unit
  - Sex
    - Attitude towards Women
    - Attitude towards Heterosexual Relationships (and Marriage)
    - Attitude towards Sexual Relationships (and Sexual Adjustment)
  - Interpersonal Relationship
    - Attitude towards Friends and Acquaintances
    - Attitude towards People Supervised
    - Attitude towards Superiors
    - Attitude towards Colleagues/Co-workers at Work or School
  - Self-Concept
    - Fears
    - Guilt Feelings
    - Attitude towards Own Abilities
    - Attitude towards Past
    - Attitude towards Future (and Goals)

Purpose in Life Tests—Part B

- **Purpose:** designed to assess an individual's sense of meaning and purpose in life
- **Population:** 13 years and older
- **Time:** usually untimed but typically takes an additional 5 to 15 minutes
- **Domains Measured:** active, emotionally stable, goal-oriented, positive direction of life
- **Consists of 3 Parts**
  - Part A: Objective Part
  - Part B: Sentence Completion
  - Part C: Essay

Draw-A-Person Test (DAPT)

- **Purpose:** to assist professionals in inferring children's cognitive developmental levels with little or no influence of other factors such as language barriers or special needs
- **Population:** 3 to 15 years 11 months
- **Time:** untimed, but individuals typically complete the drawing(s) within 10 to 15 minutes
- **Domains Measured:** measure of intellectual maturation of children, to elicit personality type and unconscious material, and as part of neuropsychological test batteries

eireandtheclouds
STUDY NOTES BY ANNE
THIS IS NOT FOR SALE
ANY FORMS OF REPRODUCTION
IS STRICTLY PROHIBITED
eireandtheclouds@gmail.com | x: rpm_anne | tiktok: eireandtheclouds

House-Tree-Person Test (HTP)

- **Purpose:** measures aspects of a person's personality through interpretation of drawings and responses to questions
- **Population:** children, adolescents, and adults
- **Time:** untimed but usually children take 150 minutes on an average to complete
- **Domains Measured:** overall psychological functioning and personality dynamics through symbolic representation

Kinetic Family Drawing Test (KFD)

- **Purpose:** a variation of traditional family drawing tests, with a unique instruction: the individual is asked to draw everyone in their family, including themselves, doing something
- **Population:** all ages
- **Time:** untimed
- **Domains Measured:** cognitive, interpersonal, or psychological functioning

Word Association Test

- **Purpose:** diagnose psychological typology and psychopathology
- **Population:** 7 and above
- **Time:** 10 to 20 minutes
- **Domains Measured:** attitudes, thoughts, desires, feelings, and even negative aspects of one's personality

Rorschach Inkblot Test

- **Purpose:** diagnosing and treating individuals with a variety of psychological problems and psychiatric disorders
- **Population:** 5 years to adult
- **Time:** untimed
- **Domains Measured:** identifying a person's state of mind and various personality traits
    - 5 black and gray inkblots
    - 2 black, gray, and red inkblots
    - 3 multicolored without black inkblots

Hand Test

- **Purpose:** to measure action tendencies, particularly action out and aggressive behavior
- **Population:** 5 years and older

Thematic Apperception Test (TAT)

- **Purpose:** to reveal an individual's perception of interpersonal relationships
- **Population:** 5 to 79 years old
- **Time:** 45 minutes to 2 hours or more
- **Domains Measured:** useful as part of a comprehensive study of personality and in the interpretation of behavior disorders, psychosomatic illness, neuroses, and psychoses
- **Versions:**
    - **Children's Apperception Test**
        - Author: Leopold Bellak and Sonya Sorel Bellak
        - Purpose: to assess personality, emotional problems, and underlying conflicts in children. It aims to uncover a child's perceptions of their world, their relationships, and their unconscious needs, motives, and anxieties through storytelling
        - Description: Consists of 10 picture cards. The original CAT (CAT-A) uses animal figures (e.g., lions, bears, monkeys) in various human-like social situations (e.g., eating at a table, playing tug-of-war, sleeping in a bed)
        - Population: 3 to 30 years old
        - Time: 20 to 45 minutes
        - Year of Publication: 1949
    - **Philippine Thematic Apperception Test**
        - Author: Dr. Alfredo V. Lagmay
        - Purpose: to elicit unconscious or preconscious material regarding an individual's personality, motivations, conflicts, and interpersonal dynamics
        - Description: consists of a set of ambiguous picture cards that depict Filipino scenes, people, clothing, social situations, and cultural nuances
        - Population: intended for adults, ages 12 and above
    - **Philippine Children's Thematic Apperception Test**
        - Author: Dr. Alfredo V. Lagmay
        - Purpose: to assess personality, emotional conflicts, and family dynamics in children
        - Description: consists of a series of picture cards featuring Filipino children and/or animals in situations common to Filipino childhood experiences
        - Population: below 12 years old

eireandtheclouds STUDY NOTES BY ANNE
THIS IS NOT FOR SALE. ANY FORMS OF REPRODUCTION IS STRICTLY PROHIBITED.

TEST ADMINISTRATION, SCORING, INTERPRETATION, & USAGE (20)

A | Detect Errors and Impacts in Test

Test Administration Procedure

- **Pre-Administration Phase**
  - **Examiner Qualification & Preparation:**
    - Master test manuals, administration scripts, and precise timing requirements.
    - Ensure appropriate credentials (e.g., Level A, B, or C user qualifications).
  - **Environment Setup:**
    - Select a quiet, well-lit, properly ventilated, and distraction-free room.
    - Arrange seating to prevent cheating and minimize physical discomfort.
    - Display "Testing in Progress — Do Not Disturb" signage.
  - **Materials Preparation:**
    - Verify all necessary test booklets, answer sheets, stopwatches, pencils, and stimulus materials are complete and secure.
  - **Informed Consent & Assent (if needed):**
    - Obtain signed consent outlining the purpose of testing, intended use of results, confidentiality limits, and fee structures.
    - Secure assent from minors or individuals lacking legal capacity.
- **Administration Phase**
  - **Rapport Building:** Establish a warm, professional, and reassuring tone to alleviate test anxiety without providing unfair encouragement or hints.
  - **Adherence to Standardized Instructions:**
    - Read verbatim instructions from the test manual.
    - Maintain precise timing using standardized timepieces where required.
  - **Observation and Behavioral Tracking:** Record non-verbal cues, frustration tolerance, attention span, and approach to task-solving.
  - **Maintaining Test Security:** Continuously monitor examinees to prevent unauthorized assistance or exposure of test items.
- **Post-Administration Phase**
  - **Data Verification and Scoring:**
    - Immediately collect and account for all test materials.
    - Check answer documents for double-marked items or incomplete demographics.
    - Apply standardized scoring keys, tables, and algorithms accurately.
  - **Debriefing and Feedback:** Provide examinees or guardians with an overview of the next steps and expected timelines for reports.
  - **Secure Storage:** Store completed protocols in locked cabinets or encrypted databases conforming to legal and ethical regulations (e.g., HIPAA, FERPA).

Issues in Test Administration

- **Examiner Effects**
  - **Expectancy/Rosenthal Effect:** Examiner expectations can subtly influence testee performance through micro-cues or biased scoring of ambiguous responses.
  - **Interpersonal Style:** Rigidly cold or overly friendly examiners can alter test-taker motivation and stress levels.
  - **Scoring Errors:** Clerical oversights, misinterpretation of subjective responses, and misuse of normative conversion tables.
- **Test Taker Variables**
  - **Test Anxiety:** High cognitive interference and physiological arousal degrade working memory and processing speed.
  - **State Factors:** Fatigue, illness, drug/medication effects, acute emotional distress, or low motivation.
  - **Test-Wiseness:** Variations in familiarity with test formats, guessing strategies, and time management skills.
- **Contextual and Environmental Factors**
  - **Physical Distractions:** Unpredictable noises, extreme temperatures, poor lighting, or cramped seating.
  - **Administrative Deviations:** Altering time limits, rephrasing instructions, or improper response prompting.
- **Accommodations vs. Modifications**
  - **Accommodations –** adjustments that minimize the impact of a disability without

eireandtheclouds STUDY NOTES BY ANNE THIS IS NOT FOR SALE. ANY FORM OF REPRODUCTION IS STRICTLY PROHIBITED

altering the underlying construct being measured (e.g., extra time for motor deficits, large-print fonts).
    - **Modifications –** alterations that change the construct being measured, potentially invalidating comparison to standardized norms (e.g., reading a reading-comprehension passage aloud).

Issues in Intelligence Testing

- **Theoretical & Conceptual Challenges**
    - Debate between unitary models (Spearman's g factor) and multidimensional models (Cattell-Horn-Carroll CHC theory, Gardner's Multiple Intelligences, Sternberg's Triarchic Theory).
    - Intelligence tests primarily measure academic aptitude and abstract reasoning, often omitting practical, emotional, or creative capabilities.
- **Score Interpretation & Misuse**
    - Treating an Intelligence Quotient (IQ) as a fixed, physical trait rather than a dynamic, sample-based statistical estimate.
    - Misusing IQ scores for rigid tracking, placement decisions, or stigmatizing diagnoses without considering adaptive functioning.
    - Ignoring confidence intervals when reporting scores
- **Generational and Temporal Shifts**
    - **The Flynn Effect:** The worldwide, historic secular rise in raw intelligence test scores (approximately 3 IQ points per decade).
- **Dynamic Assessment Alternatives**
    - Traditional static testing measures learned performance (Zone of Actual Development)
    - Dynamic assessment uses a test-intervene-retest format to evaluate learning potential (Zone of Proximal Development

Culture Bias of Testing

- **Major Types of Test Bias**
    - **Construct Bias –** the test measures different constructs across different cultural groups.
    - **Item Bias (Content) –** specific items are easier or harder for group members of equal overall ability.
    - **Predictive/Criterion Bias –** the test differential predicts criterion outcomes (e.g., GPA) for different groups.
- **Cultural Loading –** the degree to which a test requires knowledge, language, or skills unique to a specific culture.
- **Cultural Bias –** present when cultural loading leads to unfair differential performance or invalid inferences about intrinsic ability.
- **Language Proficiency –** administering tests in an examinee's non-native language measures language proficiency rather than raw cognitive capability.
- **Translation Artifacts –** direct translations often fail to preserve item difficulty, idioms, or construct equivalence across languages.
- **Strategies for Mitigating Cultural Bias**
    - **Culture-Fair & Non-Verbal Tests:** Utilizing matrix reasoning, spatial puzzles, and non-verbal instructions (e.g., Raven's Progressive Matrices, Leiter-3, UNIT-2). Note: No test is entirely culture-free.
    - **Local and Representative Norming:** Developing normative samples that adequately represent minority, linguistic, and socioeconomic demographics.
    - **Multi-Method/Multi-Source Assessment:** Combining standardized tests with naturalistic observation, adaptive behavior scales, and informant reports.
    - **Testing in Native Language:** Employing bilingual examiners or validated parallel-language instruments.

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

B | Errors in Reliability

Classical Test Theory – a score on an ability is presumed to reflect not only the test taker's true score on the ability being measured but also the **error**

- Also referred to as true score (or classical) model of measurement

TOS-BASED REVIEWERS by Anne (rpm_anne, eireandtheclouds@gmail.com) IS ENTIRELY FOR FREE | NOT FOR SALE PLEASE!

- Everyone has a "true score" on a test
- **True Score –** a value that, according to CTT, genuinely reflects an individual's ability (or trait) level as measured by a particular test
- Its assumptions allow for its application in most situations
- **Error –** the component of observed test score that does not have to do with the test taker's ability
- Errors of measurement are random
- *The greater the number of items. The higher the reliability*
    - X = T (true score) + E (error)
- 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 variables, other than variables being measured

- **Random Error –** source of error in measuring a targeted variable caused by unpredictable fluctuations and inconsistencies of other variables in the measurement process. *Unavoidable*
- **Systematic Error –** source of error in measuring a variable that is typically constant or proportionate to what is presumed to be the true value of the variable being measured. *Avoidable if corrected*

Sources of Error

- Test Construction
    - Item Sampling/Content Sampling – variation among items within a test as well as to variation among items between tests
- Test Administration
    - Test Environment (room temperature, level of lighting, and the amount of ventilation and noise)
    - Test taker variable (emotional problems, physical discomfort, lack of sleep, effects of drug, formal learning experiences, casual life experiences, therapy, illness, and changes in mood or mental state)
    - Examiner-related variables (examiner's physical appearance and demeanor, nonverbal gestures, and professionalism)

Reliability Estimate Error

- **Test Retest Reliability**
    - **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 remembered or practiced the previous test
    - **Practice Effect -** scores on the second session are higher due to their experiences of the first session of testing
    - Test-retest with longer interval might be affected of other extreme factors, thus, resulting to low correlation
    - At least 2 weeks, maximum of 6 months
- **Parallel Forms/Alternate Forms Reliability**
    - 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 test
    - Test scores may be affected by motivation, fatigue, or intervening events
- **Inter-item Consistency**
    - Item sampling homogeneity
- **Split-half Reliability**
    - Item sample: nature of split
- **Inter-rater Reliability**
    - Scorer differences

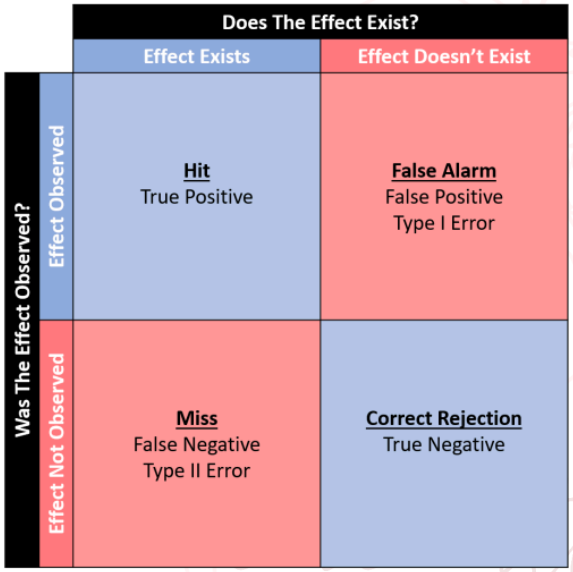
Standard Error of Measurement – provide an estimate of the amount of error inherent in an observed score of measurement

- Index of the amount of inconsistency or the amount of the expected error in an individual's score (true score)
- Allows to quantify the extent to which a test provide accurate score
- higher reliability = lower SEM
- **Standard Error of Score –** index of the extent to which one individual's score vary over tests presumed to be parallel
- **Confidence Interval –** a range or or band of the test scores that is likely to contain the true score
- **Standard Error of the Difference –** statistical measure that can aid a test user in determining a large difference should be before it is considered statistically significant
- **Standard Error of Estimate –** refers to the standard error of the difference between predicted and observed values

eireandtheclouds
TOS-BASED REVIEWERS BY ANNE
THIS IS NOT FOR SALE
ANY FORMS OF REPRODUCTION IS STRICTLY PROHIBITED

Hit and Miss Rate

- **Hit Rate –** the proportion of people a test accurately identifies as possessing or exhibiting a particular trait, behavior, characteristics, or attribute
- **Miss Rate –** the proportion of people the test fails to identify as having, or not having a particular trait, behavior, characteristics, or attribute



4 Possible Hit and Miss Outcomes

- ➔ **True Positives (Sensitivity) –** predict success that does occur
- ➔ **True Negatives (Specificity) –** predict failure that does occur
- ➔ **False Positive (Type 1) –** success that does not occur
- ➔ **False Negative (Type 2) –** predicted failure but succeed

C | Errors due to Behavioral Assessment

Observer-Related Errors

- **Observer Drift –** the gradual, unintentional shift over time in how an observer applies target behavior coding criteria (e.g., becoming more lenient or stricter after weeks of observation).
- **Observer Bias/Expectancy Effects –** when an observer's prior knowledge (e.g., knowing a child is in the intervention group) skews their recording of behaviors in expected directions.
- **Rating Scale Biases:**
  - **Halo Effect:** Allowing an overall impression of a subject to influence ratings on specific behavioral items.
  - **Leniency/Severity Error:** Consistent tendency to rate all observed behaviors as excessively favorable (leniency) or unfavorable (severity).
  - **Central Tendency Error:** Avoidance of extreme rating categories, leading to artificially clustered middle ratings.
  - **Observer Fatigue/Inattention:** Reduced recording accuracy resulting from long observation intervals or excessive cognitive strain.
- **Subject/Participant-Related Errors**
  - **Reactivity –** an alteration in a subject's natural behavior caused solely by the awareness of being observed (e.g., a student acting abnormally quiet while a supervisor is in the classroom).
  - **Demand Characteristics:** Environmental cues that inform the target individual of expected behaviors, causing them to alter performance to match the perceived expectations.
- **Methodological & Procedural Errors**
  - **Behavioral Sampling Error:**
    - **Time Sampling Error:** Observing during unrepresentative times of day (e.g., observing hyperactivity right after recess vs. during silent reading).
    - **Setting/Situational Variance:** Failing to account for environmental triggers specific to one context (e.g., home vs. school).
  - **Ambiguous Operational Definitions:** Poorly defined target behaviors (e.g., defining "aggression" without clear behavioral bounds) leading to low Inter-Observer Agreement (IOA).
  - Coding Scheme Complexity: Overly intricate coding systems with too many simultaneous categories increase human recording error and decrease reliability.

D | Sundberg's Levels of Clinical and Test Interpretation

Level I: Direct Sample/Minimal Inference

- Test data treated purely as a direct behavioral sample or direct actuarial score.
- Minimal to none. No assumption is made about hidden traits or internal psychological states.
- Example: Recording that a student typed 45 words per minute on a typing test, or utilizing automated

eireandtheclouds@gmail.com | x: rpm_anne | tiktok: eireandtheclouds
NOTES BY ANNE
NOT FOR SALE.
OF REPRODUCTION
PROHIBITED

computer-generated actuarial decision rules.

Level II: Descriptive Generalization & Signs/Intermediate Inference

- Data used as a correlate or sign of broader behavioral tendencies or underlying psychological states.
- Descriptive Generalization – expanding a specific test score into a general trait category (e.g., inferring a general trait of "high test anxiety" from an elevated score on a single self-report scale).
- Hypothetical Construct – inferring an underlying internal state or mental process (e.g., interpreting frequent handwashing on a checklist as a sign of underlying obsessive-compulsive traits).

Level III: Dynamic Case Conceptualization/High Inference

- Fully integrated, idiographic model of the individual.
- High level of holistic inference combining historical context, developmental background, environmental stressors, and multi-method assessment data.
- Formulates an inclusive theory of the person's personality structure and psychosocial functioning rather than reporting isolated test scores.
- Example: Synthesizing cognitive scores, projective data, developmental history, and clinical observations to understand how a patient's childhood attachment trauma influences current interpersonal conflicts.

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

E | Types of Interpretation

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

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

TEST 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 to make decisions
- Reliability < Validity < Utility
- Higher criterion-related validity = higher utility

Factors that Affect Utility

- **Psychometric Soundness –** the reliability and validity of a test or tool
- **Costs –** disadvantages, losses, or expenses, both economic and noneconomic
- **Benefits –** profits, gains, or advantages

Utility Analysis – it is a family of techniques that entails a cost-benefit analysis designed to yield information relevant to a decision about the usefulness and/or practical value of a tool of assessment

- It must be undertaken for the purpose of evaluating whether the benefits of using a test outweigh the costs

Conducting Utility Analysis

- ➔ **Expectancy Tables –** provide an indication that a test taker will score within some interval of scores on a criterion measure (passing, acceptable, failing)
- ➔ **Taylor-Russell 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

$$selection\ ratio = \frac{number\ of\ hired\ candidates}{total\ number\ of\ candidates}$$

- ➔ **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

➔ **Decision Theory –** developers examine the usefulness of test scores in helping the test user make decisions

Some Practical Considerations

- The pool of job applicants
- The complexity of the job
- The cut score (reference point derived as a result of a judgment and used to divide a set of data into two or more classifications) in use
    - **Relative Cut Score –** reference point based on norm-related considerations
    - **Fixed Cut Score –** set with reference to a judgement concerning minimum level of proficiency required
    - **Multiple Cut Scores –** two or more cut score with one predictor
    - **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

Methods for Setting Cut Score

- **Angoff Method –** setting fixed score, requires expert judges
- **Known Groups Method –** collection of data on the predictor of interest from groups known to possess, and not to possess, a trait, attribute, or ability of interest
- **IRT-based Methods –** cut scores are typically set based on test taker'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 test takers who have acquired the minimum 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 of acceptance, and the distribution of applicant scores
- **Discriminant Analysis –** shed lights on the relationship between identified variables and two naturally occurring groups

ETHICAL PRINCIPLES & STANDARDS OF PRACTICE (19)

Note: the following are summary only

PRINCIPLE I: RESPECT FOR THE DIGNITY OF PERSONS AND PEOPLES

- Respect for dignity recognizes the inherent worth of all human beings, regardless of perceived or real differences in social status, ethnic origin, gender capacities, or other such as characteristics
- Respect for the dignity of persons and peoples is expressed in various ways in different communities and cultures

a) Respect for the unique worth and inherent dignity
b) Respect for diversity
c) Respect for the customs and beliefs
d) Free and informed consent
e) Privacy of individuals
f) Protection of confidentiality of personal information
g) Fairness and justice
h) Respect for the environment

PRINCIPLE II: COMPETENT CARING FOR THE WELL-BEING OF PERSONS AND PEOPLE

- Involves working for their benefit and, above all, doing no harm
- Maximizing benefits, minimizing potential harm, and offsetting or correcting harm
- Requires the application of knowledge and skills that are appropriate for the nature of a situation as well as the social and cultural context
- Enhance potential benefits and reduce potential harm
- Self-knowledge of how one's values, experiences, culture, and social context might influence one's actions and interpretation
- Competent caring also covers the delivery of psychological services using online modalities (telepsychology)

a) Active concerns for the well-beings…psychological services, whether face-to-face or via online platform
b) Taking care to do no harm
c) Maximizing benefits and minimizing potential harm
d) Correcting or offsetting harmful effects…occurred as a result of their activities
e) Developing and maintaining competence
f) Self-knowledge

eireandtheclouds | STUDY NOTES BY ANNE | THIS IS NOT FOR SALE | ANY FORMS OF REPRODUCTION IS STRICTLY PROHIBITED | eireandtheclouds@gmail.com | x: rpm_anne | tiktok: @eireandtheclouds

g) Respect for the ability of individuals

PRINCIPLE III: INTEGRITY

- Advancement of scientific knowledge and to the maintenance of public confidence in the discipline of psychology
- Based on honesty, and on truthful, open and accurate communication
- The need to protect the safety or confidentiality of persons and peoples, and the need to respect cultural expectations
- Regardless of such differences, monitoring and management are needed to ensure that self-interest does not interfere with acting in the best interests of persons and peoples

a) Honest, truthful, and prudent, open and accurate communication
b) Avoiding incomplete disclosure of information unless complex disclosure is culturally inappropriate, or violates confidentiality
c) Maximizing impartiality and minimizing biases
d) Not exploiting persons or peoples
e) Avoiding conflicts of interest

PRINCIPLE IV: PROFESSIONAL AND SCIENTIFIC RESPONSIBILITIES TO SOCIETY

- Psychology functions as a discipline within the context of human society
- Contributing to the knowledge about human behavior and to persons' understanding of themselves and others
- Using such knowledge to improve the condition of individuals, families, groups, communities, and society

a) To increase scientific and professional knowledge in ways that allow the promotion of the well-being of society and all its members
b) To use psychological knowledge for beneficial purposes and to protect such knowledge from being misused, used incompetently, or made useless
c) To conduct its affairs in ways that are ethical and consistent with the promotion of the well-being of society and all its members
d) To promote the highest ethical ideals in the scientific, professional, and educational activities of its members
e) To adequately train and supervise its members in their ethical responsibilities and required competencies
f) To develop its ethical awareness and sensitivity, and to be as self-correcting as possible

GENERAL ETHICAL STANDARDS AND PROCEDURES

A Resolving Ethical Issues

Misuse of the Psychology Professional's Works

- We take appropriate and reasonable steps to correct or minimize the effects of such misuse or misrepresentation

Conflicts between Ethics and Law, Regulations or other Governing Legal Authority

- Our first step is to take appropriate actions to resolve the conflicts while being committed to our Code of Ethics

Informal Resolution of Ethical Violation

- Resolve the issue by bringing it to the attention of the concerned professional
- We do so if informal resolution is sufficient and if the intervention does not violate confidentiality rights

Reporting Ethical Violations

- We take further action to report violation of the Code of Ethics to appropriate institutional authorities
- Does not apply when an intervention would violate confidentiality rights...called to review the work of another psychology practitioner

Cooperating with Ethics Committee

- Cooperate with the ethics investigation

Improper Complaints

- Refrain from filing ethical complaints with reckless disregard or willful ignorance of facts that would disprove allegations of ethical violations
- Refrain from filing complaints without supporting factual evidence

Unfair Discrimination Against Complainants and Respondents

- We do not discriminate against complainants and respondents of ethical complaints by denying them employment, advancement, admission to academic, tenure or promotion

STUDY NOTES BY ANNE
THIS IS NOT FOR SALE
ANY FORMS OF REPRODUCTION IS STRICTLY PROHIBITED
x: rpm_anne | tiktok: eireandtheclouds

PSYCHOLOGICAL ASSESSMENT

PmLE | **Main References: Cohen & Swerdlik, Kaplan & Sacuzzo, Miller & Lovler, Urbina**

B Competencies

RA 10029 (The Philippine Psychology Act of 2009) – only licensed and registered psychologists and psychometricians are legally allowed to practice their profession

C Human Relations

Unfair Discrimination

- We shall not discriminate against persons based on age, gender, gender identity, race, ethnicity, culture, national origin, regional identity, religion, sexual orientation, exceptionality, occupation, socioeconomic status, educational background, or any basis proscribed by law

Sexual Harassment

- We do not engage in sexual harassment (*RA 7877 or Philippine Anti-Sexual Harassment*)

Other Harassment

- We shall take reasonable steps to avoid harming of our clients/patients

Avoiding Harm

- We shall take reasonable steps to avoid harming

Multiple Relationships

- Refrain from entering into a multiple relationship…reasonably be expected to impair our objectivity, competence, or effectiveness in performing our functions as psychologists or psychometricians, or otherwise risks exploitation or harm to the person with whom our professional relationship exists

(1) Is in another role with the same person
(2) is in a relationship with a person closely associated with or related to the person with whom the psychology practitioner has the professional relationship
(3) Promises to enter into a future relationship with that client/patient or a person closely associated with or related to that client/patient

D Confidentiality

Maintaining Confidentiality

- Safeguard any information divulged by our client, regardless of the medium where it stored

Limitations of Confidentiality

- Duty to discuss the limitations of confidentiality to our clients, may it due to regulated laws, institutional rules, or professional or scientific relationship
- Minor or is legally incapable clients: the primary guardian or legal representative should be informed about the limitations of confidentiality (assent form: minor/incapable clients)
- Explain explicitly to the client all anticipated uses of the information they will disclose
- If the psychological services, products, or information are coursed through an electronic transmission it is our duty to inform the clients of risks to privacy

Disclosures

- Information to be disclosed will not be misused, misunderstood or misinterpreted to infringe on human rights, whether intentionally or unintentionally
- Disclose confidential information only when the client or legal representative gave their consent, unless it is prohibited by law
- Disclose confidential information only to the source of referral and with a written permission from the client if it is self-referral
- Disclose confidential information without the consent of the client or legal representative only when it is mandated by law or permitted by law
- We divulge information is limited to what is only needed to achieve the purpose
- We make sure that only qualified psychology practitioners will supervise such releases
- Release of such information becomes confidential and may be acquired only with prior approval from management or an authorized representative

ETHICAL STANDARDS AND PROCEDURES—ASSESSMENT

Bases for Assessment

- We provide through our recommendations, reports, and diagnostic or evaluative statements are based on substantial information and appropriate assessment techniques
- Only after employing adequate assessment procedures and examination to support our conclusions and recommendations
- We discuss the limitations of our opinions and the basis of our conclusions and recommendations
- Can be used as basis for making decisions about clients or test takers as long as these are administered by qualified professionals, following

eireandtheclouds STUDY NOTES BY ANNE THIS IS NOT FOR SALE ANY FORMS OF REPRODUCTION IS STRICTLY PROHIBITED x: rpm_anne | tiktok: eireandtheclouds

prescribed assessment standards

Informed Consent in Assessment

- Gather informed consent prior to the assessment
  - When it is mandated by the law
  - When it is implied, such as in routine educational, institutional and organizational activity
  - When the purpose of the assessment is to determine the individual's decisional capacity
- Educate our clients about the nature of our services, financial arrangements, potential risks, and limits of confidentiality…not competent to provide informed consent on assessment, we discuss these matters with immediate family members or legal guardians
- The confidentiality of test results and the security of the tests must be ensured
- We inform test takers of the implications of poor internet connectivity during the test-taking session on their test performance, particularly for timed assessments

Assessment Tools

- Judiciously select and administer only those tests which are pertinent to the reasons for referral and purpose of the assessment
- Use data collection, methods and procedures that are consistent with current scientific and professional developments
- We use tests that are standardized, valid, reliable, and has a normative data directly referable to the population of our clients
- Appropriate to the language, competence and other relevant characteristics of our client
- We use only original companies of test materials (not photocopies)

Obsolete and Outdated Test Results

- Do not base our interpretations, conclusions, and recommendations on outdated test results
- Do not provide interpretations, conclusions, and recommendations on the basis of obsolete tests

Interpreting Assessment Results

- We should therefore indicate our reservations regarding the interpretations
- Interpret assessment results while considering the purpose of the assessment and other factors

Release of Test Data

- Our responsibility to ensure that test results and interpretations are not used by persons other than those explicitly agreed upon by the referral sources
- We do not release test data in the forms of raw and scaled scores

Explaining Assessment and Results

- Only to the sources of referral and with a written permission from the client if it is a self-referral
- Explain findings and test results to be communicated to relatives, parents, or teachers…we explain them through a non-technical language
- Except when the relationship precludes the provision of explanation of results and it is explained in advance to the client
- When test results needs to be shared with schools, social agencies, the courts or industry, we supervise such releases
- Clients may request from the psychologist or psychometrician if the session can be recorded

Test Security

- Shall be handled only by a qualified/licensed users or personnel

Assessment by Unqualified Persons

- We do not promote the use of assessment tools and methods by unqualified persons except for training purposes with adequate supervision
- Test protocols…their interpretations and all other records are kept secured from unqualified persons
- As psychometricians, we practice under the boundaries prescribed in RA10029. We operate under the supervision of a registered psychologist, when applicable

Test Construction

- Using current scientific findings and knowledge, appropriate psychometric properties, validation, and standardization procedures

Other Psychology-Related Laws

- → **RA 9258** Guidance and Counseling Act of 2004
- → **RA 9262** Violence Against Women and Children
- → **RA 7610** Child Abuse
- → **RA 7877** Philippine Anti-Sexual Harassment
- → **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
- → **RA 11650** Inclusive Education Law
- → **RA 10173** Data Privacy Act
- → **RA 12080** Basic Education Mental Health and Well-Being Promotion Act

STUDY NOTES BY ANNE | NOT FOR SALE | ANY FORMS OF REPRODUCTION IS STRICTLY PROHIBITED.

- **HB 4982** SOGIE Bill
- **HB 9611** An Act Amending Republic Act No. 10029, Otherwise Known as the "Philippine Psychology Act of 2009" and for Other Purposes

ADDITIONAL INFORMATION: RA10029 (PHILIPPINE PSYCHOLOGY ACT OF 2009)

- **Psychology –** is the scientific study of human behavior. It involves the application of scientific methods to inquire into the biological, cognitive, affective, developmental, personality, social, cultural and individual difference dimensions of human behavior

THE MAJOR ROLE OF REGISTERED PSYCHOMETRICIAN

- Administering and scoring of objective personality tests, structured personality tests, excluding tests and other level forms of psychological tests
- Interpreting results of the same and preparing a written report on these results
- Conducting preparatory intake interviews of clients for psychological intervention sessions

Ratings in the Examination

- To be qualified as having passed the licensure examination for psychologists and psychometricians, a candidate must have obtained a weighted general average of **at least seventy-five percent (75%) for all subjects**, with **no grade lower than sixty percent (60%) in any given subject**
- An examiner who **obtains a weighted general average of seventy-five percent (75%) or higher but obtains a rating below sixty percent (60%) in any given subject** may **retake such subjects within the next two (2) years**, and upon obtaining a rating of at least seventy-five percent (75%) in each such subject, shall then be deemed to have passed the licensure examination

RA11036 (MENTAL HEALTH ACT) — OBJECTIVES

- Strengthen effective leadership and governance for mental health by, among others, formulating, developing, and implementing national policies, strategies, programs, and regulations relating to mental health
- Develop and establish a comprehensive, integrated effective and efficient national mental health care system responsive to the psychiatric, neurologic, and psychosocial needs of the Filipino people
- Protect the rights and freedoms of persons with psychiatric, neurologic, and psychosocial needs; Filipino people
- Strengthen information systems, evidence and research for mental health
- Integrated mental health care in the basic health services
- Integrate strategies promoting mental health in educational institutions, the workplace, and in communities

MENTAL STATUS EXAMINATION – is a structured assessment used to evaluate a person's current mental state

- It involves observing and documenting various aspects of a person's cognitive, emotional, and behavioral functioning at a specific point in time
- Appropriate before assessment to determine the appropriateness of more formal psychological testing

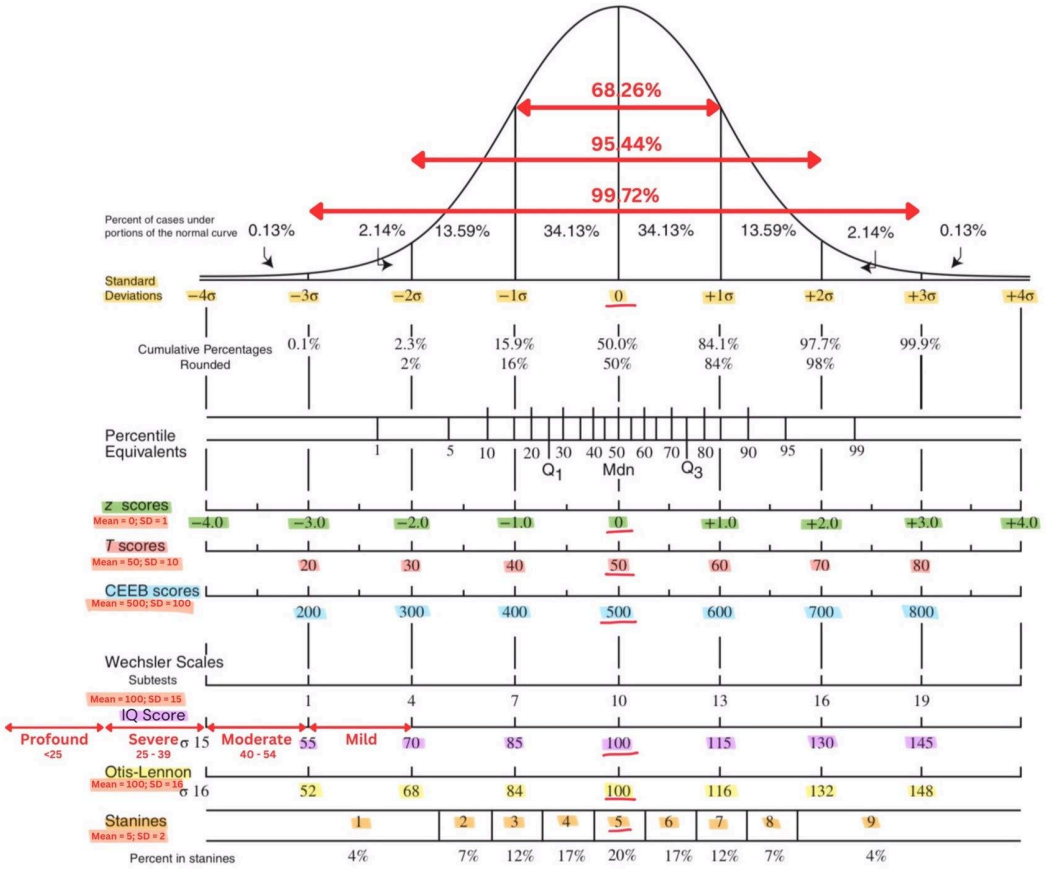
7 Key Domains of MSE

- **Appearance and Behavior –** observes the patient's overall physical presentation and conduct
- **Speech –** evaluates the patient's verbal communication
- **Mood and Affect**
  - **Mood –** is the patient's subjective, self-reported emotional state (e.g., "I feel sad")
  - **Affect –** is the clinician's objective observation of the patient's emotional expression (e.g., "The patient has a flat affect")
- **Thought Process and Content**
  - Thought Process refers to the form or how a person thinks
  - Thought Content refers to the substance or what a person is thinking about
- **Perception –** specifically looks for perceptual disturbances like hallucinations or illusions
- **Cognition**
  - **Orientation –** awareness of person, place, and time
  - **Attention and Concentration –** the ability to focus and sustain attention
  - **Memory –** immediate, recent, and remote memory
  - **Intellectual Functioning –** an assessment of general knowledge and abstract thinking
- **Insight and Judgement**
  - **Insight –** is the patient's understanding

of their own illness, including its nature and severity

- **Judgment –** is the patient's ability to make sound decisions and act appropriately in various situations

END



Disclaimer: I got this from rpmtwt but I forgot who is the OP. credits to the rightful owner

eireandtheclouds@gmail.com | x: rpm_anne | tiktok: eireandtheclouds