



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPLE

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

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



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

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

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:

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

R_{xx} = correlation coefficient

x = obtained score

$\bar{x}$ = mean

Goals of Reliability

- Estimate errors in psychological measurement
- Develop techniques to improve testing so errors are reduced

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

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** –



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

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)

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 – obtained by correlating pairs of scores from the sample people on two different administrations

- **Coefficient of Stability** – the longer the time that passess, the greater the likelihood that the reliability coefficient will be lower
- **Pearson Product-Moment Correlation Coefficient (r)** – is typically calculated between the scores from the two administrations
- **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** – item sampling and other errors have affected scores on versions of the same test
 - **Parallel Forms** – each form of the test means that the variance of observed test scores is equal
- **Alternate Forms Reliability** – an estimate of the extent to which these different forms of the same test have been affected by item sampling error, or other errors
 - **Alternate Forms** – simple, different versions of the test that have been constructed so as to be parallel
- **Counterbalancing** – a technique to avoid carryover effects for parallel forms, by using different sequence for groups
- Two different but rigorously constructed equivalent forms of the test are administered to the same group of individuals.
- The forms should be similar in content, difficulty, and format
- Pearson product-moment correlation coefficient (r) is calculated between the scores on the two parallel forms



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

Internal Consistency (Inter-Item Consistency) –

degree of correlation among all the items on a scale

- **Homogeneity** – single factor test measure
- **Heterogeneity** – multiple factor test measure
- **Kuder-Richardson Formula 20** – used exclusively for tests where items are dichotomously scored (i.e., items with only two possible responses, such as correct/incorrect, true/false, or yes/no)
 - Different degrees of difficulty
- **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)
- **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

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

- **Spearman-Brown Formula** – estimate internal consistency reliability from a correlation of two halves of a test
- **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



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

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** – lower correlation coefficient
- **Inflated** – higher correlation coefficient

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

B Reliability Models and Theories

Classical Test Theory

- Also referred to as 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

- 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

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



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

over tests presumed to be parallel

→ **Confidence Interval** – a range or band of the test scores that is likely to contain the true score

- ◆ 68% confidence interval (± 1 SEM); this means there's a 68% chance that the individual's true score falls within that range
- ◆ 95% confidence interval (± 1.96 SEM; often approximated as ± 2 SEM); this means there's a 95% chance that the individual's true score falls within that range

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 - $(1 \times \text{SEM}) = 75 - (1 \times 3) = 75 - 3 = 72$
- ◆ Upper bound: Observed score + $(1 \times \text{SEM}) = 75 + (1 \times 3) = 75 + 3 = 78$
- ◆ 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 - $(2 \times \text{SEM}) = 75 - (2 \times 3) = 75 - 6 = 69$
- ◆ Upper bound: Observed score + $(2 \times \text{SEM}) = 75 + (2 \times 3) = 75 + 6 = 81$
- ◆ 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

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

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



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

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
- **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



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPLP

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

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)** – can 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 criterion
 - Often used when a new test is



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

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



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

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 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



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

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

	Method 1			Method 2			Method 3		
Traits	Trait 1	Trait 2	Trait 3	Trait 1	Trait 2	Trait 3	Trait 1	Trait 2	Trait 3
Method 1	(.90)								
	.10	(.90)							
	.10	.10	(.90)						
Method 2	.50	.12	.12	(.80)					
	.14	.50	.12	.20	(.80)				
	.14	.14	.50	.20	.20	(.80)			
Method 3	.50	.05	.05	.40	.03	.03	(.85)		
	.04	.50	.05	.02	.40	.03	.30	(.85)	
	.04	.04	.50	.02	.02	.40	.30	.30	(.85)

Notes: The correlations are artificial. Reliabilities are in parentheses. Heterotrait-monomethod correlations are in the gray subdiagonals. Heterotrait-heteromethod correlations are enclosed by a broken line. Monotrait-heteromethod correlations in the convergent validity diagonals are in bold type.

Eid, M. (2010). Multitrait-multimethod matrix. In Encyclopedia of research design (Vol. 0, pp. 851-855). SAGE Publications, Inc., <https://doi.org/10.4135/9781412961288.n255>

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

- **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



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

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 from 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



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

→ **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

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

Criterion-Referenced

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
$\alpha \geq 0.90$	Excellent
$0.80 \leq \alpha < 0.90$	Good
$0.70 \leq \alpha < 0.80$	Acceptable
$0.60 \leq \alpha < 0.70$	Questionable
$0.50 \leq \alpha < 0.60$	Poor
$\alpha < 0.50$	Unacceptable



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

Interrater Reliability Index

K	1.0	Excellent		Almost Perfect	Excellent
	0.9				
	0.8				
	0.7	Good	Fair to Good	Substantial	Very 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	

Cicchetti & Sparrow, 1981

Fleiss, 1981

Landis & Koch, 1977

Regier et al. 2012 - DSM-5

K	1.0	Excellent	Excellent	Almost Perfect	(Excellent)
	0.9				
	0.8				
	0.7	Good	Fair to Good	Substantial	Very 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	

Cicchetti & Sparrow, 1981

Fleiss, 1981

Landis & Koch (1977)

Regier et al. (2012) - DSM-5

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

p-Value

- $p\text{-value} \leq \alpha$, reject null hypothesis
- $p\text{-value} > \alpha$, accept null hypothesis

$p\text{-value} > 0.10$	No evidence against H_0
$0.05 < p\text{-value} \leq 0.10$	Moderate evidence against H_0
$0.01 < p\text{-value} \leq 0.05$	Strong evidence against H_0
$0.001 < p\text{-value} \leq 0.01$	Very strong evidence against H_0
$p\text{-value} \leq 0.001$	Extremely strong evidence against H_0



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

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
- ◆ 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



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

- **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

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

- 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

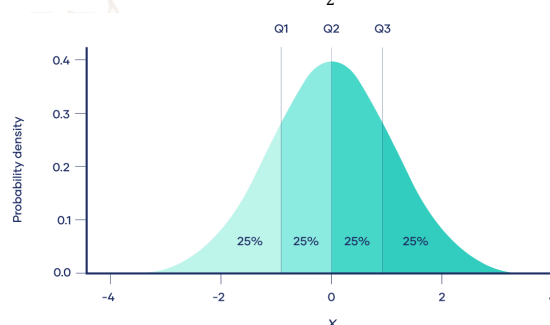
$$\text{range} = \text{highest score} - \text{lowest score}$$

- **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

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

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

- **Percentage** – means of comparing quantities

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

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

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

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

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

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

- Univariate: -3 to +3



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

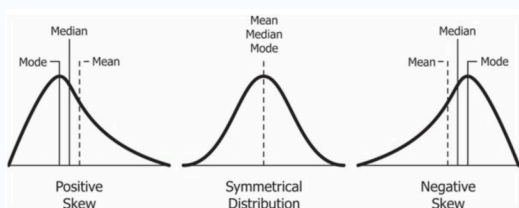
Types of Skewness

→ **Positive Skew** – relatively few of the scores fall at the high end of the distribution (difficult test)

◆ $\text{mean} > \text{median} > \text{mode}$

→ **Negative Skew** – relatively few of the scores fall at the low end of the distribution (easy test)

◆ $\text{mean} < \text{median} < \text{mode}$



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

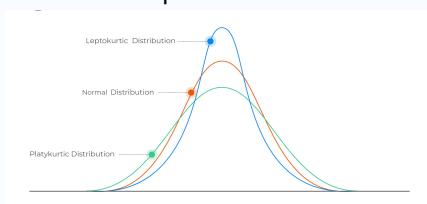
◆ Thin-tailed, meaning that outliers are infrequent

→ **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



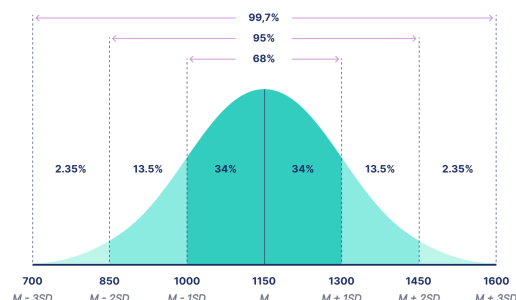
	Category		
	Mesokurtic	Platykurtic	Leptokurtic
Tailedness	Medium-tailed	Thin-tailed	Fat-tailed
Outlier frequency	Medium	Low	High
Kurtosis	Moderate (3)	Low (< 3)	High (> 3)
Excess kurtosis	0	Negative	Positive
Example distribution	Normal	Uniform	Laplace

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 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





PSYCHOLOGICAL ASSESSMENT

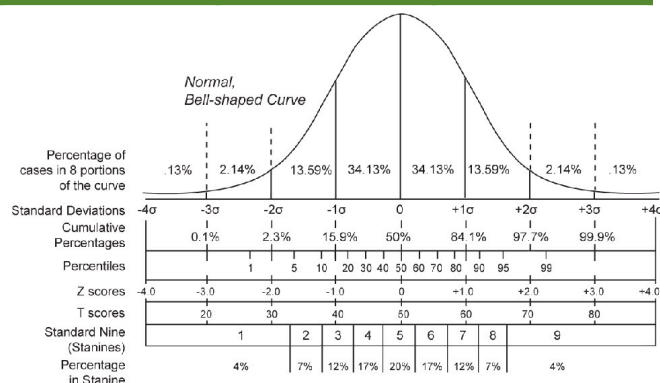
BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

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

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

B

Inferential Statistics

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

Hypothesis Testing – an educated guess that sets the direction 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”

	Parametric	Non-parametric
Assumed Distribution	Normal	Any
Assumed Variance	Homogenous	Heterogenous
Typical Data	Interval/Ratio	Nominal/Ordinal
Data Set Relationship	Independent	Any
Statistical Tool	Pearson r Independent &	Spearman rho Wilcoxon



PSYCHOLOGICAL ASSESSMENT

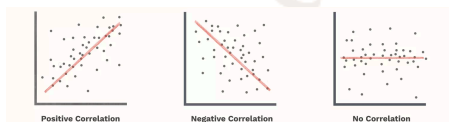
BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

	Dependent T-Test Repeated Measures ANOVA One-/Two-way ANOVA	Signed Mann Whitney U Test Friedman Test Kruskal-Wallis Test
--	---	---

Measure of Bivariate Correlation – connection between 2 variables

- **Correlation** – the degree and direction of correspondence between two things
 - Positive: $\uparrow\uparrow$ or $\downarrow\downarrow$
 - Negative: $\uparrow\downarrow$ or $\downarrow\uparrow$
 - No Correlation: —



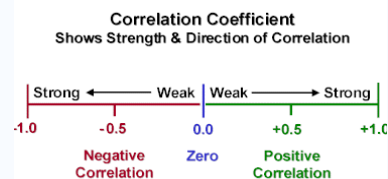
- **Coefficient of Determination (r^2)** – analyzes how the differences in first variable can be explained by a difference in a second variable
- **Coefficient of Alienation** – proportion of variance in the DV that is not accounted for by the IV
- **True Dichotomy** – naturally divides things into two 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

Pearson r	interval/ratio + interval/ratio
Spearman's Rho	ordinal + ordinal
Biserial	art. dichotomous +

	interval/ratio
Point Biserial	true dich. + interval/ratio
Phi Coefficient	nominal (true dich) + nominal (true/art. dich)
Tetrachoric	art. dich + art. dich
Kendall's	3 or more ordinal/rank
Rank Biserial	nominal + ordinal

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
 - $p > 0.05$ = 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 to Fit** – determines if the observed



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

frequencies of a single categorical variable match an expected distribution

- **Test Independence** – is used to determine if two categorical variables are related

	Goodness to Fit	Test of Independence
What is Measured	Difference	Difference + Relationship
Type of Data	Nominal	Nominal
No. of Variables	1 variable with ≥ 2 levels	2 variables with ≥ 2 categories

Comparison of 2 Groups

- **Paired Groups** – measures 1 group with 2 data sets
 - BP before and after lecture
- **Unpaired Groups** – measures 2 different groups
 - BP of male and female

Parametric Test		Non-parametric Test
Independent T-Test	2 separate groups, random assignment	Mann Whitney U Test
Dependent T-Test	1 group, two scores	Wilcoxon Signed Rank Test

Comparison of 2 or More Groups

- **Match Groups** – measures 1 group with ≥ 2 data sets
- **Unmatched Groups** – measures ≥ 3 groups with different data sets

Parametric Test		Non-parametric Test
Repeated Measures ANOVA	1 group, measured at least 3 times	Friedman Rank Test
One-Way ANOVA	3 or more groups, tested once	Kruskal-Wallis Test
Two-Way ANOVA	3 or more groups, tested for 2 variables	
ANCOVA	Blend of ANOVA and regression; IV (covariate) and DV	
ANOVA Mixed Design	2 or more groups, measured more than 3 times	
MANOVA	1 IV $\geq$ 1 DV	

Post Hoc Test – test used after an ANOVA to determine which groups are significantly different from each other

- **Tukey's Honestly Significant Difference (HSD) Test** – equal variance and sample size
- **Scheffé's Test** – more conservative and flexible test for all possible simple and complex contrasts

Regression – used when one wants to provide framework of prediction of the basis of one factor in order to predict the probable value of another factor



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPLE

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

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

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

Kolmogorov-Smirnov Test	More than 50 samples
Shapiro-Wilk Test	Less than 50 samples

Levene's Test – is a statistical test used to assess the equality of variances for a variable calculated for two or more groups

→ **P-value > Significance Level (e.g., 0.05)**

◆ If the p-value is greater than your chosen significance level (commonly 0.05), you **fail to reject the null hypothesis**

◆ There is no statistically significant difference between the variances, and you can assume homogeneity of variance

→ **P-value ≤ Significance Level (e.g., 0.05)**

◆ If the p-value is less than or equal to your chosen significance level, you **reject the null hypothesis**

◆ There is a significant difference between the variances, meaning the assumption of homogeneity of variance is violated

Bartlett's Test – used to check if multiple samples come from populations with equal variances

→ **Null Hypothesis (H_0):** population variances are equal across all groups

→ **Alternative Hypothesis (H_1):** at least one group has a different variance

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?



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

- 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



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

→ **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 – process of setting rules for assigning numbers in measurement

- **Notion of Absolute Scaling** – procedure obtaining a measure of item difficulty across samples of test takers who vary in ability

Types of Scaling

- Age-based
- Grade-based
- Stanine
- Unidimensional
- Multidimensional

Comparative Scaling

Paired Comparison	Produces ordinal data by presenting with pairs of two stimuli, which they are asked to compare
Rank Order	Respondents are presented with several items simultaneously and asked to rank them in order or priority
Constant Sum	Respondents are asked to allocate a constant sum of units, such as points, among set of stimulus objects with

	respect to some criterion
Q-Sort Technique	Sort objects based on similarity with respect to some criterion

Non-Comparative Scaling

Continuous Rating	Rate the objects by placing a mark at the appropriate position on a continuous line that runs from one extreme of the criterion variable to the other
Itemized Rating	Having numbers or brief descriptions associated with each category
Likert Scale	Indicate their own attitudes by checking how strongly they agree or disagree with carefully worded statements that range from very positive to very negative towards attitudinal object
Visual Analogue Scale	A 100-mm line that allows subjects to express the magnitude of an experience or belief
Semantic Differential Scale	Measures an individual's attitudes or emotional reactions to a concept or object by using a series of bipolar adjectives or phrases
Stapel Scale	Uses a numerical scale, typically presented vertically, with a single adjective in the middle to measure respondents' attitudes or opinions about a specific subject
Summative	Summing all the rating across



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

Scale	
	all the items
Thurstone Scale	Uses a set of statements about a topic, with each statement assigned a numerical value reflecting the respondent's attitude towards it
Ipsative Scale	Where respondents distribute a fixed number of points across different attributes, highlighting their relative strengths and weaknesses within themselves

Scoring Items

- **Class Scoring (Category Scoring)** – test taker responses earn credit toward placement in a particular class or category with other test takers whose pattern or responses is presumably similar in one way
- **Cumulative Model** – the higher the score on the test, the higher the test taker is on the ability, trait, or other characteristic that the test purports to measure
- **Ipsative Scoring** – comparing a test taker's score on one scale with a test to another scale within that same test

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

Guidelines of Writing Items

- Define clearly what to measure
- Generate item pool
- Avoid long items
- Keep the level of reading difficulty appropriate for those who will complete the test
- Avoid double-barreled items
 - ◆ Double-Barreled Items – convey more than one ideas at the same time
- Consider making positive and negative word items

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

- Denoted as “p”



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPLP

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

- 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 Endorsement Index – statistics provides not a measure 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

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

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

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



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

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



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

- **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 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



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

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	Preliminary data collection Needed in order to formulate initial hypothesis or to gain information on client's history
Psychodiagnostic Interview (used by Psychiatrists)	Questions are typically anchored on the diagnostic criteria (DSM) to assess client and provide diagnosis
Behavioral Interview (used by Psychologists)	Questions typically explore the onset of behavior Focuses on antecedents or consequences of behavior to find relevant information and know what aspects of behavior can be modified by therapy
Mental Status Examination	A structured assessment of a person's current mental capacity, encompassing their appearance, behavior, cognitive functions, and emotional state
Suicide Assessment Interview	Makes use of rating scales to assess or detect suicidal thoughts

Documents

- External source of verification
- **Case History Profile** – a comprehensive



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

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 variables, 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 observing and recording a person's behavior in their natural environment or a simulated setting
Self-Monitoring	Individuals systematically observe and record their own behavior, thoughts, or feelings
Analogue Studies	Published studies that investigate certain behaviors in a controlled environment (prototype or references)
Situational Analysis Measure	Subjects who are clients in certain situations (adults)
Role Play	The assessee is directed to act as if they were

	somewhere in a particular situation (children)
Psychophysiological Method	Measuring physiological responses that are associated with psychological states or behaviors

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 reconstruction 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



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

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 domain yield different rather than confirmatory results

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

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

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)

	Ability	Typical Performance
What is Measured	What client can do	What client typically does
Conditions	Where client is expected to perform	Anytime, anywhere
Nature of Answers	Right or wrong	All answers are correct
Objective of Motivation	Correct answers as possible	Honest and self-aware



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

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
- **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



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

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 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



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPLE

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

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



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

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 bias 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 Immediate feedback Identifies gaps and skills Reduces “faking good” Engaging and experiential	Artificiality Performance anxiety Subjectivity in evaluation 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



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

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 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

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

- 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



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

INDIVIDUAL TEST ADMINISTRATION (VERBAL & NONVERBAL TESTS)

Latest Stanford-Binet Intelligence Test (SB-5) [C]

- **Purpose:** designed for the testing of cognitive abilities
- **Population:** 2 to 85+ years
- **Score:** Verbal, Performance, and Full Scale Scores.
- **Time:** approximately 5 minutes for 10 subtests, varies depending on age and functional level of examinee (most items are unlimited, adding to overall time for administration)
 - Full Scale IQ: 45-75 minutes
 - V & NV Scales: 30 minutes each
 - Abbreviated Battery IQ: (2 routing subtests): 15-20 minutes
 - Individual Scales: 15-75 minutes
- **Domains Measured:**
 - **(FR) Fluid Reasoning** – the ability to solve verbal and nonverbal problems using inductive and deductive reasoning
 - **(K) Knowledge** – a person's accumulated fund of general information acquired at home, school, or work
 - **(QR) Quantitative Reasoning** – an individual's facility with numbers and numerical problem-solving, whether with word problem
 - **(VsP) Visuo-Spatial Processing** – measures an individual's ability to see patterns, relationships, spatial orientations, or gestalt among diverse pieces of a visual display
 - **(WM) Working Memory** – a class of memory processes in which diverse information stored in short-term memory is inspected, sorted, or transformed

- **Scales:** Full Scale IQ (10 subtests), Domain Scores (V & NV IQ), 5 Factor Indices (FR/K/QR/VsP/WM)
- **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

- First published intelligence test to provide organized and detailed administration and scoring instructions
- **Lewis Terman (1877-1956)** – translated in English the Binet-Simon test of at Stanford University, and the first American test to employ the concept of IQ
- **Alternate Item** – an item to be substituted for a regular item under specified conditions
- **Maud Merrill (1888-1978)** – development of two equivalent forms innovated in 1937, labeled L (Lewis) and M (Maud), as well as new types of tasks for use with preschool-level and adult-level test takers
- **Mental Age** – the age levels at which an individual aspect to be functioning intellectually as indicated by the level of items responded to correctly
- **Ratio IQ** – the ratio of the test taker's mental age divided by chronological age, multiplied by 100
$$\text{ratio IQ} = \frac{\text{mental age}}{\text{chronological age}} \times 100$$
- **SB-III** – deviation IQ was used in place of ratio IQ
 - ◆ **Deviation IQ** – reflects a comparison of the performance of the individual with the performance



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

of others of the same age in the standardization sample

→ SB:FE (4th ed.)

- ◆ **Point Scale** – a test organized into subsets by categorizing at which most test takers are presumed capable of responding in the way that is keyed as correct
- ◆ **Test Composite** – a test score or index derived from the combination of, and/or a mathematical transformation of, one or more subtest scores

→ SB-5

- ◆ Based on the CHC theory of intellectual abilities
- ◆ Measures 5 CHC factors by different types of tasks and subtests as different levels
- ◆ **Routing Test** – a task used to direct or route the examinee to a particular level of questions
- ◆ 2 routing tests of the SB5
 - Object series
 - Matrices and vocabulary
- ◆ **Teaching Items** – designed to illustrate the task required and assure the examiner that examinee understands
- ◆ **Floor** – lowest level of the items on a subtest
- ◆ **Ceiling** – highest level of the items on a subtest
- ◆ **Basal Level** – a subtest with reference to a specific taker's performance
- ◆ **Adaptive Testing** – testing individually tailored to the test taker at the beginning a subtest with a question on the middle range of difficulty

Wechsler Adult Intelligence (WAIS-V) [C]

- **Purpose:** to provide a comprehensive measure of an individual's intellectual functioning

- Diagnose intellectual disabilities or learning disorders
- Identify intellectual giftedness
- Assess cognitive strengths and weaknesses
- Help in neuropsychological evaluations, such as for traumatic brain injury or dementia

- **Population:** 16 years 0 months to 90 years 11 months

• Score:

- **Full-Scale IQ (FSIQ)** – a composite score that represents overall intellectual ability
- **Primary Index Scores** – five key scores that measure cognitive abilities in distinct domains
- **Subtest Scaled Scores** – individual scores for each subtest, with a mean of 10 and a standard deviation of 3
- **Ancillary and Complementary Index Scores** – additional scores that provide more specific clinical insights, such as nonverbal or quantitative reasoning abilities

- **Time:** (45 - 60) minutes

• Domains Measured:

- **Verbal Comprehension Index (VCI)** – ability to use verbal reasoning and acquired knowledge
- **Visual-Spatial Index (VSI)** – capacity to analyze, synthesize, and mentally manipulate visual information
- **Fluid Reasoning Index (FRI)** – ability to solve novel problems and identify underlying patterns
- **Working Memory Index (WMI)** – ability to hold and process information in short-term memory
- **Processing Speed Index (PSI)** – quick and efficient processing of simple visual information

• Subtests:

- **Verbal Comprehension:** Similarities, Vocabulary
- **Visual-Spatial:** Block Design, Visual Puzzles



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Sacuzzo, Miller & Lovler

- **Fluid Reasoning:** Matrix Reasoning, Figure Weights
- **Working Memory:** Digit Span, Picture Span
- **Processing Speed:** Coding, Symbol Search

Wechsler Adult Intelligence (WAIS-IV) [C]

- **Purpose:** used to assess an individual's intellectual functioning and cognitive strengths and weaknesses
 - Assisting in the diagnosis of intellectual disabilities and learning disorders
 - Providing insights into cognitive function following a brain injury, stroke, or in the context of degenerative conditions like dementia
 - Helping to identify an individual's intellectual capabilities for academic or career guidance
- **Population:** 16 years 0 months to 90 years 11 months
- **Score:**
 - **Full-Scale IQ (FSIQ)** – a single composite score that represents a person's overall intellectual ability
 - **Primary Index Scores** – four main index scores that represent functioning in specific cognitive domains
 - **Subtest Scaled Scores** – scores for each individual subtest, with a mean of 10 and a standard deviation of 3
- **Time:** (60 - 90) minutes
- **Domains Measured:**
 - **Verbal Comprehension Index (VCI)** – measures verbal reasoning, comprehension, and the ability to use and express verbal information. It reflects acquired knowledge and verbal concept formation.
 - **Perceptual Reasoning Index (PRI)** – assesses non-verbal reasoning,

visual-spatial processing, and visual-motor integration. It involves solving abstract, non-verbal problems and interpreting visual information.

- **Working Memory Index (WMI)** – measures the ability to temporarily hold and mentally manipulate information to complete a task. It's crucial for higher-order thinking and problem-solving.
- **Processing Speed Index (PSI)** – evaluates the ability to quickly and accurately process simple visual information, as well as cognitive efficiency and fluency.

- **Scales:** Full Scale IQ (4 main scales and their subtests)
 - **VCI:** Similarities, Vocabulary, Information; *Comprehension*
 - **PRI:** Block Design, Matrix Reasoning, Information; *Picture Completion, Figure Weights*
 - **WMI:** Digit Span, Arithmetic; *Letter-Number, Sequencing*
 - **PSI:** Symbol Search, Coding; *Cancellation*

Notes: FSIQ—significantly influenced by WMI, PSI, VCI, PRI; General Ability Index (GAI)—significantly influenced by VCI and PRI; Cognitive Proficiency Index (CPI)—significantly influenced by WMI and PSI

- **Core and Supplemental Subtests:**
 - **Similarities** – client is asked to identify the qualitative relationship between pair of words
 - **Vocabulary** – requires the client to try define up to 30 words
 - **Information** – aimed at determining how much general knowledge the client has accumulated from their environment
 - **Comprehension** – the individual is asked questions about social and other situations



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

- **Block Design** – consists of two-dimensional designs which the client tries to copy using three-dimensional blocks
- **Matrix Reasoning** – a nonverbal reasoning task in which individuals are asked to identify patterns in design
- **Picture Completion** – the examinee views a picture with an important part missing and identifies the missing part
- **Figure Weights** – an examinee views scales with missing weights and selects the response option that is best suited to keep the scales balanced
- **Digit Span** – individual tries to repeat digits forward, individual tries to repeat digits backward, and individual tries to repeat digits in ascending order
- **Arithmetic** – consists of a series of mental arithmetic questions
- **Letter-Number Sequencing** – the individual is read a series of letters and numbers and is required to repeat them back with the letters in alphabetical order and the numbers in numerical order
- **Visual Puzzles** – individuals view a completed puzzle and then select three response options that can be combined to reconstruct the puzzle
- **Symbol Search** – the client, under pressure, scans a search group and indicates whether one of the symbols in the target group matches
- **Coding** – individuals are asked to record associations between different symbols and numbers within time limits
- **Cancellation** – requires that subjects scan a structured arrangement of colored shapes and mark the targets and avoid the distractors

Other Wechsler Tests

→ Wechsler Intelligence Scale for Children (WISC-V)

- ◆ 6 to 16 years old
- ◆ 65–80 minutes (depending on the number of primary and secondary subtests given)
- ◆ Administered individually, one-on-one by a trained psychologist, which allows for a personalized assessment of a child's abilities.

→ Wechsler Preschool and Primary Scales of Intelligence (WPPSI-IV)

- ◆ 2–6 to 3–11 years old
- ◆ 30 to 45 minutes (ages 2:6 to 3:11); 45 to 60 minutes (ages 4:0 to 7:7); assessment varies depending on the child's age, attention span, and individual circumstances
- ◆ The WPPSI assessment results are interpreted by comparing the child's scores to the standardized norms based on age and other relevant factors. The scores are typically reported as scaled scores, percentile ranks, and composite scores
- ◆ There are three distinct age bands each child may fall into when being scored
 - Age band 1 – 0 to 3 year
 - Age band 2 – 4 to 7 year
 - Age band 3 – 8 to 11 year

GROUP TEST ADMINISTRATION (NONVERBAL TESTS)

Raven Progressive Matrices (RPM) [B]

- **Purpose:** designed to measure a person's ability to form perceptual relations
- **Population:** 4 to 90 years old
- **Score:** Percentile Ranks
- **Time:** 40 to 60 minute
- **Domains Measured:** Abstract Reasoning



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

Versions of Raven Progressive Matrices

→ Standard Progressive Matrices (SPM)

- ◆ **Purpose:** nonverbal assessment tool designed to measure an individual's ability to perceive and think clearly, make meaning out of confusion, and formulate new concepts when faced with novel information
- ◆ **Population:** 6 to 16 years and adults of average intelligence (8 to 65 years old), general population
- ◆ **Score:** Percentile Ranks
- ◆ **Time:** 40 to 45 minutes
- ◆ **Description:** The Standard Progressive Matrices (SPM) was designed to measure a person's ability to form perceptual relations and to reason by analogy independent of language and formal schooling, and may be used with persons ranging in age from 6 years to adult

→ Advanced Progressive Matrices (APM)

- ◆ **Purpose:** measures high-level observation skills, clear thinking ability, and intellectual capacity as nonverbal estimate of abstract reasoning or fluid intelligence
- ◆ **Population:** 11 years old and above, specifically gifted students
- ◆ **Time:** 40 to 60 minutes

→ Colored Progressive Matrices – used to assess the degree to which children and adults with moderate to severe learning difficulties

- Scale 3: 14 to college students and adults of superior intelligence
- **Time:**
 - Test 1: Series – 3 minutes
 - Test 2: Classification – 4 minutes
 - Test 3: Matrices – 3 minutes
 - Test 4: Topology (2.5 minutes)
- **Domains Measured:**
 - **Series** – the individual is presented with incomplete, progressive matrices; their task is to select, from among the choices provided, the answer that best continues the series
 - **Classifications** – the individual is presented with 5 figures. In scale 2, you must select one that is different from the other four. In scale 3, you must correctly identify two figures which are in some way different from 3 others
 - **Matrices** – the task is to correctly complete the design or matrix presented at the left each row
 - **Conditions (Topology)** – requires the individual to select, from the five choices provided, the one which duplicates the conditions given in the far left box

Purdue Non-Language Test (PNLT) [B]

- **Purpose:** measure individual intelligence in manner designed to reduce, as much as possible, the influence of verbal fluency, cultural climate, and educational level
- **Population:** 13 years old and above
- **Time:** 25 minutes
- **Domain Measured:** Abstract Reasoning

Panukat ng Katalinuhang Pilipino (PKP) by Palacio

- **Purpose:** designed to measure the mental ability of a Filipino adult on the basis of his verbal and nonverbal skills
- **Population:** 16 years old and above
- **Time:** 52 minutes
- **Domains Measured:**
 - Verbal skills
 - Talasalitaan (Vocabulary)

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

- **Purpose:** designed to assess intelligence without being hindered by cultural and environmental factors
- **Population:**
 - Scale 1: 4 to 8 years and older, mentally handicapped individuals
 - Scale 2: 8 to 14 years old and average adults



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Sacuzzo, Miller & Lovler

- Ugnayan (Analogy)
- Nonverbal skills
 - Kakayahan sa mga Bilang (Numerical Ability)
 - Isinalarawan Problema (Abstract Reasoning)

Differential Aptitude Tests (DAT-5)

- **Purpose:** measures your abilities and skills, such as verbal and numerical ability, abstract reasoning, mechanical reasoning and space relations
- **Population:** Grades 7 to 12 years old and adults
- **Time:** 3 hours (timed for each battery)
- **Domain Measured:**
 - **Verbal Reasoning** – measures the ability of a student to see relationships among words
 - **Numerical Ability** – measures the ability to perform mathematical reasoning tasks
 - **Abstract Reasoning** – assesses how well individuals can reasoning geometric shapes and designs
 - **Perceptual Speed and Accuracy** – measures the ability to compare and make written lists quickly and accurately
 - **Mechanical Reasoning** – measures the ability to understand basic mechanical principles of machinery, tools, and motion
 - **Space Relation** – measures the ability to visualize a three-dimensional object from a two-dimensional pattern and to visualize how this object would look if rotated in space
 - **Spelling** – measures how well the student can spell common English words
 - **Language Use** – measures the ability to detect errors in grammar, punctuation, and capitalization
- **Versions:**
 - **DAT Form R** – (Grades 7 to 10: Standard Form)

- **DAT Form S** – (Grades 7 to 10: Advanced Form)
- **DAT Form K** – (Grades 10 to 12 and Adults: Standard Form)
- **DAT Form L** – (Grades 10 to 12 and Adults: Advanced Form)

Scales:

- Educational Aptitude = Verbal Reasoning + Numerical Reasoning
- General Intelligence = Verbal Reasoning + Numerical Reasoning + Abstract Reasoning

Thurstone Tests of Mental Alertness

- **Purpose:** evaluates the general mental ability and capability to learn new information and acquire new skills through 126 numerical and verbal questions
- **Population:** literate adolescent and adults
- **Time:** 20 minutes (less than 10 seconds for each question)
- **Scoring:**
 - L Score of linguistic or verbal ability
 - Q Score of quantitative ability
 - Total score is the sum of L score and the Q score

Wonderlic Cognitive Ability Test (WCAT)

- **Purpose:** measure cognitive ability and problem-solving aptitude of prospective employees for a range of occupations
- **Population:** 15 years old and up
- **Time:** administration of the paper-and-pencil version takes about 20 minutes. The exact time is 12 minutes; the other 8 minutes are for administrative procedures

Watson Glaser Critical Thinking Tests (WGCT)

- **Purpose:** designed to measure ability to reason through an argument logically and make an objective decision
- **Population:** 16 to 64 years old
- **Time:** 30 to 60 minutes

Flanagan Industrial Tests (FIT)

- **Purpose:** used for selection, placement, reassignment, and vocational counseling and



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

help employers select potential employees

- **Population:** college-bound high school juniors and seniors and adults in personnel selection programs
- **Time:** 5 to 15 minutes
- **Domains Measured:**
 - **Arithmetic** – ability to add, subtract, multiply and divide
 - **Assembly** – ability to visualize how separate pieces will look as a whole
 - **Components** – ability to visualize how separate pieces will look as a whole
 - **Coordination** – ability to control hand and arm movements while working through a series of mazes
 - **Electronics** – ability to understand electrical and electronic principles as well as analyze diagrams of electrical circuits
 - **Expression** – knowledge of correct grammar and sentence structure
 - **Ingenuity** – ability to think of ingenious and effective ways of solving problems
 - **Inspection** – ability to spot imperfections or flaws in a series of objects
 - **Judgment & Comprehension** – ability to read and comprehend given information
 - **Mathematics & Reasoning** – ability to reason through mathematical word problems
 - **Mechanics** – ability to understand mechanical principles and analyze mechanical movement
 - **Memory** – ability to memorize different terms and their meanings
 - **Patterns** – ability to perceive and reproduce pattern outlines accurately
 - **Planning** – ability to plan, organize and schedule various types of activities
 - **Precision** – capacity for precision work with small objects
 - **Scales** – ability to read scales, graphs, and charts

- **Tables** – ability to read tables quickly and accurately
- **Vocabulary** – knowledge of words used in business and government environments

Philippine Aptitude Classification Test (PACT)

- **Purpose:** identifies specific abilities of high school students for the purpose of educational and vocational guidance; attempts to predict a student's probable performance in various courses of study
- **Population:** Grade 9 (3rd-year student), ages 14 to 15
- **Time:** Part 1: 4 minutes, Part 2: 1 hour and 45 minutes
- **Domains Measured:**
 - Perceptual Speed
 - Verbal English
 - General Reasoning
 - Flexibility of Closure
 - Verbal Filipino
 - Spatial Closure
 - Visualization
 - Perceptual Acuity

B

Personality Inventory Tests/Projective Tests

Basic Personality Inventory [C]

- **Purpose:** is a personality assessment intended for use with clinical and normal populations to identify sources of maladjustment, consisting 240 true-false items
- **Population:** 12 years old to adults
- **Time:** 35 minutes
- **Domains Measured:**
 - **Alienation** – antisocial/socially deviant attitudes
 - **Anxiety** – general anxiety, panic reaction
 - **Denial** – response style, lack of insight
 - **Depression** – depressive affect and hopelessness
 - **Deviation** – often considered a



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Sacuzzo, Miller & Lovler

critical item scale measuring critical deviant behaviors

- **Hypochondriasis** – preoccupation with bodily concerns
- **Impulse Expression** – inability to consider consequences of actions
- **Interpersonal Problems** – difficulties cooperating with others
- **Persecutory Ideas** – paranoia and perception of external control
- **Self-Depreciation** – negative self-perception
- **Social Introversion** – avoidance of social interactions due to disinterest, fear, or withdrawal
- **Thinking Disorder** – disorganized thinking, problems with reality testing
- **Higher-Order Factors**
 - **Measures of Mood & Personal/Emotional Adjustment:** Hypochondriasis, Depression, Anxiety
 - **Measures of Inadequate or Deviant Socialization & Impulse:** Interpersonal Problems, Alienation, Impulse Expression
 - **Measures of Cognitive Functioning:** Persecutory Ideas, Thinking Disorder
 - **Measures of Self-Perception & Sociability:** Self-Depreciation, Social Introversion
 - **Measure of Critical Deviant Behaviors:** Deviation
 - **Measure of Test-Taking Style:** Denial

Myers-Briggs Type Indicator (MBTI)

- **Purpose:** designed to help people understand personality differences in the general population
- **Population:** 13 years and older
- **Time:** no time limit
- **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:** consists of 225 statements that came from daily life. To provide researchers in career development with a quick and convenient measure of a number of relatively independent personality variables, or needs



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

- **Population:** 16 to 85 years old
- **Score:** 15 scores
- **Time:** 45 minutes
- **Domains Measured:**
 - **Achievement (Ach)** – a need to accomplish tasks well
 - **Deference (Def)** – a need to conform to customs and defer to others
 - **Order (Ord)** – a need to plan well and be organized
 - **Exhibition (Exh)** – a need to be the center of attention in a group
 - **Autonomy (Aut)** – a need to be free of responsibilities and obligations
 - **Affiliation (Aff)** – a need to form strong friendships and attachments
 - **Intracception (Int)** – a need to analyze behaviors and feelings of others
 - **Succorance (Suc)** – a need to receive support and attention from others
 - **Dominance (Dom)** – a need to be a leader and influence others
 - **Abasement (Abe)** – a need to accept blame for problems and confess errors to others
 - **Nurturance (Nur)** – a need to be of assistance of others
 - **Change (Cha)** – a need to seek new experience and avoid routine
 - **Endurance (End)** – a need to follow through on tasks and complete assignments
 - **Heterosexuality (Het)** – a need to be associated with and attractive to members of the opposite sex
 - **Aggression (Agg)** – a need to express one's opinion and be critical of others
- **Validity Scales:**
 - EPPS has used the last 15 items to offer the candidate the same item twice, using the results to calculate a consistency score. The results **will be considered valid** if the consistency checks for more than **9 out of 15 paired items**

- **Purpose:** designed to measure personality and can also be used by psychologists, and other mental health professionals, as a clinical instrument to help diagnose psychiatric disorders, and help with prognosis and therapy planning
- **Population:** 16 years and above
- **Score:** STEN scores
- **Time:** 30 to 50 minutes
- **Domains Measured:**
 - **Primary Factors** – basic definition of individual personality difference; more detailed information about the richness and uniqueness of the individual
 - **Global Factors** – second-order; higher and more theoretical definitions of the personality; provide a broad overview of personality

Global Factors	Primary Factors
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

Sixteen Personality Factor Questionnaire (16PF)



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

Self-Control/Lack of Restraint	(F) Lively-Serious (G) Rule-conscious/Expedient (M) Abstracted-Practical (Q3) Perfectionistic-Tolerates disorder
--------------------------------	---

• Validity Scales:

- **Impression Management (IM) Scale** – a bipolar scale with high scores reflecting a preponderance of socially desirable responses and low scores reflecting preponderance of socially undesirable responses
- **Acquiescence (ACQ) Scale** – indexes the degree to which the examinee agreed with items regardless of what was being asked
- **Infrequency (INF) Scale** – comprises the most statistically infrequent responses on the test, which are all middle responses and appear in the test booklet with a question mark

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

- **Purpose:** designed for use in employment and personal counseling settings involving activities such as career counseling, career development, and employee training, where these four domains are the main focus
- **Population:** 17 to 89 years old
- **Score:** 30 facet scores and 5 domain scores
- **Time:** 30 to 40 minutes
- **Domains Measured:**
 - **Openness to Experience** – the active seeking and appreciation of experiences for their own sake
 - **(O1) Fantasy** – a vivid imagination and an active fantasy life
 - **(O2) Aesthetics** – deep appreciation for art, beauty, and creative expression
 - **(O3) Feelings** – openness to

one's own feelings and emotions, both positive and negative

- **(O4) Actions** – preference for novelty and variety; willingness to try new things
- **(O5) Ideas** – intellectual curiosity and a preference for abstract ideas and philosophical discussions
- **(O6) Values** – openness to re-examining traditional values and beliefs; intellectual liberalism
- **Conscientiousness** – assess the degree of organization, persistence, control, motivation in goal-directed behavior
 - **(C1) Competence** – feels capable, effective, and able to handle situations
 - **(C2) Order** – neat, tidy, and organized
 - **(C3) Dutifulness** – strong sense of duty and obligation; adheres to rules and principles
 - **(C4) Achievement Striving** – works hard to achieve goals; ambitious and driven
 - **(C5) Self-discipline** – ability to motivate oneself to complete tasks; persistent
 - **(C6) Deliberation** – cautious and thoughtful; tends to think things through before acting
- **Extraversion** – the quantity and intensity of preferred interpersonal interactions, activity level, need for stimulation, and capacity for joy
 - **(E1) Warmth** – friendly, affectionate, and compassionate towards others
 - **(E2) Gregariousness** – preference for being with people; enjoys large groups



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

- and social gatherings
- **(E3) Assertiveness** – dominant, forceful, and outspoken
- **(E4) Activity** – energetic, rapid personal tempo, active and busy
- **(E5) Excitement-Seeking** – craves excitement and stimulation; enjoys thrills and adventurous activities
- **(E6) Positive Emotions** – tendency to experience joy, happiness, love, and enthusiasm
- **Agreeableness** – the kinds of interactions an individual prefers along a continuum from compassion to antagonism
 - **(A1) Trust** – believes in the honesty and good intentions of others
 - **(A2) Straightforwardness** – frank, sincere, and honest in interactions
 - **(A3) Altruism** – actively concerned with the welfare of others; willing to help
 - **(A4) Compliance** – yielding and cooperative; avoids conflict
 - **(A5) Modesty** – humble and unassuming; not boastful.
 - **(A6) Tender-Mindedness** – sympathetic and concerned for others; empathetic
- **Neuroticism** – the chronic level of emotional adjustment and instability
 - **(N1) Anxiety** – prone to worry, apprehension, nervousness
 - **(N2) Angry Hostility** – tendency to experience anger, frustration, and bitterness
 - **(N3) Depression** – tendency to experience sadness, guilt, hopelessness, and loneliness

- **(N4) Self-Consciousness** – sensitivity to ridicule, feelings of inferiority, shyness, and social anxiety
- **(N5) Impulsiveness** – difficulty controlling cravings and urges, leading to behaviors that may later be regretted
- **(N6) Vulnerability** – inability to cope with stress, feeling dependent, hopeless, or panicked in difficult situations

- **Validity Scales:**

- **(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
- **Versions:** Original- 1985 (NEO-PI); Revised- 1992 (NEO-PI-R); 3rd Edition- 2002 (NEO-PI-3)
 - **Form S** – self-report format
 - **Form R** – administered by an observer

Pictorial Self-Concept Scale for Children by Pasao

- **Purpose:** the Pasao Self-Concept Rating Scale, which is bilingual, evolved from the need for a self-concept scale applicable to Philippine culture. It is applicable to the high school group and can be used for counseling, general guidance, and research
- **Population:** children in kindergarten to high school
- **Time:** 25 to 40 minutes

Panukat ng Ugali at Pagkatao (PUP by Enriquez)

- **Purpose:** designed to measure Filipino-oriented traits, behaviors, and attitudes, primarily to identify incentive talent



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Sacuzzo, Miller & Lovler

- **Population:** no exact age range stated
- **Time:** 30 to 45 minutes
- **Domains Measured:**
 - **Extraversion/Surgency**
 - (+) Ambisyon (Ambition)
 - (+) Lakas ng Loob (Guts/Daring)
 - (-) Pagkasunud-sunuran (Conformity)
 - (-) Pagkamahiyain (Shyness/Timidity)
 - **Agreeableness**
 - (+) Pagkamapagkumbaba (Humility)
 - (+) Pagkamatulungin (Helpfulness)
 - (+) Pagkamaagbigay (Generosity)
 - (+) Pagkamagalang (Respectfulness)
 - (-) Pagkamapunahin (Criticalness)
 - (-) Pagkapalaaway (Belligerence)
 - (-) Hirap Kausapin (Difficulty to Deal w/)
 - **Conscientiousness**
 - (+) Katiyahaan (Perseverance)
 - (+) Pagkaresponsible (Responsibleness)
 - (+) Kapatiran (Thriftiness)
 - (+/-) Pagkasigurista (Prudence)
 - (-) Pagkasalawahan (Fickle-mindedness)
 - (-) Tigas ng Ulo (Stubbornness)
 - **Emotional Stability**
 - (+/-) Pagkamapagtimpi (Restraint)
 - (-) Pagkapikon (Low Tolerance for Teasing)
 - (-) Pagkamaramdamin (Sensitiveness)
 - (-) Sumpung (Mood)
 - **Intellect/Openness to Experience**

- (+) Pagkamausisa (Inquisitiveness)
- (+) Pagkamaalalahanin (Thoughtfulness)
- (+) Pagkamalikha (Creativity)

- **Validity Scales:**

- **Pagkakaila** (deny basic truths)
- **Kaugalian** (to reject cultural values typically held by Filipinos)
- Note: when a test taker's responses on item belonging to the 2 scales indicate these response tendencies, his/her responses on the rest of the items are no longer analyzed or interpreted

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:** 13 years old and above
- **Time:**
 - **Combined 3 Forms** – about 50 minutes to 1 hour (max. 1 hour and 30 minutes)
 - **Longer forms (Form K, Form S)** – 20 to 30 minutes
 - **Shorter form (Form KS)** – 10 to 20 minutes
- **Domains Measured:**
 - **Form K/ Form K (Interpersonal Scales)** – personality dimensions relevant to interpersonal relationships
 - Pagkamaalalahanin (Thoughtfulness)
 - Pagkamadaldal (Social Curiosity/Talkativeness)
 - Pagkamagalang



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Sacuzzo, Miller & Lovler

- (Respectfulness)
- Pagkamaramdamin (Sensitiveness)
- Pagkamasunurin (Obedience)
- Pagkamatulungin (Helpfulness)
- Pagkamaunawain (Capacity to be Understanding)
- Pagkapalakaibigan (Sociability/Friendliness)
- **Porma S/ Form S (Personal Scales) –** personality traits
 - Pagkamaayos/Orderliness
 - Pagkamahinahon/Emotional Stability
 - Pagkamapagkumbaba/Humility
 - Pagkamasayahin/Cheerfulness
 - Pagkamatapad/Honesty
 - Pagkamatiyaga/Patience
 - Pagkaresponsible/Responsibility
- **Porma KS/ Form KS (Intelligence/Creativity Scales) –** intelligence-creativity
 - Pagkamalikha/Creativity
 - Pagkamapagsapalaran/Risk Taking
 - Pagkamasikap/Achievement Orientation
 - Pagkamatalino/Intelligence

C

Interest Tests

Strong Campbell Interest Inventory Test

- **Purpose:** provides robust insight into person's interests, so you can help them to consider potential careers, their educational path and the world of work
- **Population:** 13 years old and up, reading level 9th grade
- **Time:** 35 to 40 minutes
- **Scoring:** report consists of four scores:

administrative index and special scales, general occupational themes, basic interest scale, occupational scale

- **Based on Holland's RIASEC Model**

- **Realistic (R):** Individuals who prefer working with their hands, tools, machines, or outdoors (e.g., engineers, mechanics, farmers)
- **Investigative (I):** Individuals who enjoy analyzing, researching, and solving problems (e.g., scientists, researchers, doctors)
- **Artistic (A):** Individuals who value self-expression, creativity, and aesthetics (e.g., artists, musicians, writers, designers)
- **Social (S):** Individuals who enjoy helping, teaching, or caring for others. (e.g., teachers, counselors, nurses, social workers)
- **Enterprising (E):** Individuals who are persuasive, adventurous, and enjoy leading or influencing others (e.g., entrepreneurs, sales managers, politicians)
- **Conventional (C):** Individuals who prefer working with data, numbers, and well-defined procedures (e.g., accountants, administrative assistants, librarians)

Thurstone Interest Schedule

- **Purpose:** a checklist by which a person can systematically clarify their understanding of their vocational interest
- **Population:** Adults
- **Time:** requires less than 10 minutes of subject's time

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)

- **Purpose:** assess and monitor the reading, spelling, and math skills, and helps identify possible learning disabilities



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Sacuzzo, Miller & Lovler

- **Population:** 5 to 85+ years old
- **Time:**
 - 5 to 7 years old – 15 to 25 minutes
 - 8+ years old – 35 to 45 minutes
- **Domains Measured:**
 - **Word Reading** – assesses untimed letter identification and word recognition. It involves the examinee reading aloud a list of letters and words (decoding skills)
 - **Sentence Comprehension** – measures the ability to understand the meaning of words and to comprehend the ideas and information presented in a sentence using an untimed modified cloze procedure (reading comprehension)
 - **Spelling** – measures an individual's ability to write letters and words from dictation without a time limit (orthographic knowledge and phonetic skills)
 - **Math Computation** – measures an individual's ability to count, identify numbers, solve simple oral math problems, and perform written math calculations within a time limit (arithmetic, algebra, geometry, and advanced operations)
- **Scales:**
 - Overall Composite
 - Reading Composite = Word Reading + Sentence Comprehension

B. Intelligence Tests

SRA Verbal and Nonverbal Tests

- **Purpose:**
 - SRA Verbal – to measure an individual's adaptability and flexibility in problem solving
 - SRA Nonverbal – to measure general intelligence independent of English language knowledge
- **Population:**
 - SRA Verbal – persons who are familiar with the English language
 - SRA Nonverbal – persons with reading

difficulties, illiterates and foreign-born

- **Time:**
 - SRA Verbal – 15 minutes
 - SRA Nonverbal – 10 minutes
- **Domains Measured (SRA Verbal only):**
 - Vocabulary (Linguistic)
 - Arithmetic Reasoning (Quantitative)

Test of Nonverbal Intelligence (TONI-4)

- **Purpose:** measure general intelligence, aptitude, abstract reasoning, and problem-solving abilities in a nonverbal, motor-reduced format
- **Population:** 6 years 0 months to 89 years 11 months
- **Time:** 15 to 20 minutes
- **Domains Measured:** abstract reasoning and figural problem-solving

Kaufman Assessment Battery for Children-II (KABC-II)

- **Purpose:** serves multiple purposes in psychoeducational, clinical, and neuropsychological evaluations
- **Population:** 3 years 0 months to 18 years 11 months
- **Time:**
 - Core Battery (Luria model): approx. 25 to 55 minutes
 - Core Battery (CHC model): approx. 30 to 75 minutes
 - Extended Battery: up to 100 minutes
- **Domains Measured:**
 - **(Gsm) Sequential Processing** – ability to process information in a step-by-step, serial, or linear manner
 - **(Gv) Simultaneous Processing** – ability to organize and integrate information into a whole, often involving visual-spatial reasoning and problem-solving
 - **(Glr) Learning Ability** – ability to learn and recall new information, often through associative memory
 - **(Gf) Planning Ability** – ability to solve novel problems and use mental operations such as drawing



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

inferences and applying inductive or deductive reasoning

Woodcock-Johnson IV Tests (WJ-IV)

- **Purpose:** used to evaluate cognitive abilities, academic achievement, and oral language proficiency
- **Population:** 2 years 0 months to 90+ years
- **Time:**
 - 5 to 10 minutes per subtest
 - 30 to 75 minutes per battery
- **Domains Measured:** aspects of 7 different broad CHC abilities
 - (Gf) Fluid Intelligence
 - (Gv) Visual-Spatial Ability
 - (Gs) Processing Speed
 - (Glr) Long-Term Retrieval
 - (Ga) Auditory Processing
 - (Gsm) Short-Term Memory
 - (GC) Crystallized Intelligence
- **Test Batteries:**
 - **Woodcock-Johnson IV Tests of Cognitive Abilities (WJ IV COG)** – used to identify learning problems and individual strengths and weaknesses
 - **Woodcock-Johnson IV Tests of Achievement (WJ IV ACH)** – a test battery used to measure math and reading proficiency and compare academic achievement in relation to subject's academic knowledge
 - **Woodcock-Johnson IV Tests of Oral Language (WJ IV OL)** – a test battery used to assess language proficiency

C. Aptitude tests

Otis-Lennon School Ability Test (OLSAT-8)

- **Purpose:** designed to measure a student's cognitive abilities that are most closely related to success in school
- **Population:** kindergarten to grade 12
- **Time:** 60 to 75 minutes
- **Domains Measured:**
 - Verbal Comprehension
 - Following Directions
 - Antonyms

- Sentence Completion
- Verbal Reasoning
 - Aural Reasoning
 - Arithmetic Reasoning
 - Logical Selection
 - Word/Letter Matrix
 - Verbal Analogy
 - Verbal Classification
 - Sentence Arrangement
- Pictorial Reasoning
 - Picture Classification
 - Picture Analogies
 - Picture Series
- Figural Reasoning
 - Figural Classification
 - Figural Analogies
 - Pattern Matrix
 - Figural Series
- Quantitative Reasoning
 - Number Series
 - Number Inference
 - Number Matrix

• 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)

- **Purpose:** designed to describe how individuals see themselves and how others perceive them in everyday life
- **Population:** 13 years and older
- **Time:** 45 to 60 minutes
- **Domains Measured:**
 - Dealing with Others
 - (Do) Dominance
 - (Cs) Capacity for Status
 - (Sy) Sociability
 - (Sa) Self-Acceptance
 - (In) Independence
 - (Em) Empathy
 - Self-Management



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

- (Re) Responsibility
- (So) Socialization
- (Sc) Self-Control
- (Gi) Good Impression
- (Wb) Sense of Well-Being
- (To) Tolerance
- Motivations and Thinking Style
 - (Ac) Achievement via Conformance
 - (Ai) Achievement via Independence
 - (Ie) Intellectual Efficiency
- Personal Characteristics
 - (Py) Psychological-Mindedness
 - (Fx) Flexibility
 - (F/M) Femininity/Masculinity

E. Psychopathology Personality Tests

Minnesota Multiphasic Personality Inventory II (MMPI-2)

- **Purpose:** to provide a standardized, objective measure of an individual's personality structure and symptoms of psychopathology
 - Diagnose mental disorders and assist with treatment planning
 - Evaluate individuals in forensic, public safety, and employment contexts, such as for pre-employment screening for law enforcement
 - Assess the effectiveness of therapy over time by comparing scores from multiple administrations
- **Score:** T-score; T-score of 65 or higher is considered clinically significant, indicating a deviation from the norm that may reflect psychopathology
- **Population:** aged 18 years and older
- **Time:** 567 true-false questions; 60 to 90 minutes
- **Clinical Scales**
 - **(Hs) Scale 1: Hypochondriasis** – bodily functions, physical symptoms, and health
 - **(D) Scale 2: Depression** – symptoms of depression, pessimism, low morale, dissatisfaction

- **(Hy) Scale 3: Hysteria** – physical symptoms under stress, denial of psychological problems
- **(Pd) Scale 4: Psychopathic Deviate** – antisocial behavior, rebelliousness, family problems, difficulties with authority
- **(Mf) Scale 5: Masculinity-Femininity** – traditional masculine or feminine interests and roles (less used for psychopathology, more for vocational/personality exploration)
- **(Pa) Scale 6: Paranoia** – suspiciousness, feelings of persecution, delusional thinking
- **(Pt) Scale 7: Psychasthenia** – anxiety, obsessive-compulsive tendencies, phobias, difficulties concentrating
- **(Sc) Scale 8: Schizophrenia** – unusual thought processes, social alienation, bizarre experiences, poor impulse control
- **(Ma) Scale 9: Hypomania** – elevated mood, excessive energy, impulsivity, grandiosity
- **(Si) Scale 0: Social Introversion** – withdraw from social contact, shyness, discomfort in groups

● **Validity Scales:**

- **(CNS) Cannot Say** – questions not answered (left blank or both true and false)
- **(L) Lie Scale** – attempts to present oneself in an overly favorable or virtuous light, often by denying minor flaws
- **(F) Infrequency Scale** – infrequent or atypical responses, which can suggest exaggeration of problems (faking bad), random responding, or genuine severe psychopathology
- **(K) Correction Scale** – defensiveness and a tendency to present oneself in an overly positive light (faking good)
- **(VRIN) Variable Response Inconsistency** – inconsistent answering patterns



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

- **(TRIN) True Response Inconsistency** – tendency to answer “true” or “false” indiscriminately, regardless of content (acquiescence or non-acquiescence)
 - **(F-K) F minus K Scale** – honesty of test responses
 - **(S) Superlative Scale** – exaggeratedly virtuous presentation
 - **(Fp) Infrequent Psychopathology** – overreporting of rare symptoms not typically associated with genuine psychopathology, useful in forensic context
 - **(FBS) “Faking Bad Scale”/Symptom Validity** – over-reporting symptoms in individuals with psychopathology
 - **(RBS) Response Bias Scale** – exaggerated memory complaints in forensic settings or disability claims
 - **(Fs) Infrequent Somatic Response** – overreporting of somatic symptoms
 - **(CRIN) Combined Response Inconsistency** – combination of random and fixed inconsistent responding
 - **Versions:**
 - MMPI (original version)
 - MMPI-A – for adolescents aged 14 to 18
 - MMPI-2-RF – a shorter (338 items) and psychometrically streamlined version of the MMPI-2
 - MMPI-A-RF – adolescent version of the Restructured Form
 - MMPI-3 – published in 2020, features updated items, scales, and norms, and is also shorter (335 items), often taking 25-50 minutes to complete
- Minnesota Multiphasic Personality Inventory III (MMPI-3)**
- **Purpose:** to assist clinicians in the diagnosis and assessment of major psychological disorders
 - Evaluate an individual's psychological state
 - Identify potential personality disorders or psychopathology
 - Inform treatment planning and management decisions
 - Provide an empirical basis for expert testimony in legal contexts
 - **Population:** 18 years and older
 - **Score:** T-score; T-score of 65 or higher is typically considered a clinically significant elevation, suggesting the presence of psychopathology or an exaggeration of symptoms
 - **Time:**
 - Computerized administration typically takes 25 to 35 minutes
 - Paper-and-pencil administration generally takes 35 to 50 minutes
 - **Clinical Scales**
 - **(H-O) Higher-Order Scales** – provide a broad overview of a test taker's psychological functioning
 - **(EID) Emotional/Internalizing Dysfunction** – measures problems related to mood, anxiety, and internalizing distress
 - **(THD) Thought Dysfunction** – assesses disorganized and psychotic thinking
 - **(BXD) Behavioral/Externalizing Dysfunction** – reflects issues with impulse control, aggression, and antisocial behavior
 - **(RC) Restructured Clinical Scales** – were developed to address item overlap the original MMPI clinical scales; each RC scale measures a specific, distinct clinical construct
 - **(RCd) Demoralization** – general distress, unhappiness, and dissatisfaction
 - **(RC1) Somatic Complaints** – preoccupation with physical health problems
 - **(RC2) Low Positive Emotions**



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

- lack of enjoyment, anhedonia, and low morale
 - **(RC3) Cynicism** – belief that others are malevolent and exploitative
 - **(RC4) Antisocial Behavior** – measures antisocial attitudes and behaviors
 - **(RC6) Ideas of Persecution** – paranoia and persecutory ideation
 - **(RC7) Dysfunctional Negative Emotions** – anxiety, irritability, and worry
 - **(RC8) Aberrant Experiences** – unusual thought processes and perceptual distortions
 - **(RC9) Hypomanic Activation** – energetic, impulsive, and manic-like behavior
- **Validity Scales**
 - **(?) Cannot Say** – a count of items left unanswered; too many omitted items can invalidate the result
 - **(VRIN) Variable Response Inconsistency** – detects random or inconsistent responding
 - **(TRIN) True Response Inconsistency** – identifies a tendency to answer all items as either “true” or “false”
 - **(F, Fp, Fs) Infrequent Responses** – measure a tendency to overreport (faking bad) psychopathology
 - **F** – infrequent responses that are rarely endorsed by the general population
 - **Fp** – infrequent responses specific to psychopathology, helping to distinguish between genuine severe psychopathology and malingering
 - **Fs** – infrequent responses related to somatic concerns
 - **(L-r) Uncommon Virtues** – measures a deliberate attempt to present oneself in an unrealistically positive

light

- **(K-r) Adjustment Validity** – more subtle measure of defensiveness, identifying a tendency to deny or minimize psychological problems
- **(RBS) Response Bias Scale** – measures a tendency to overreport memory or cognitive complaints

Millon Clinical Multiaxial Inventory 3rd Edition (MCMI-III)

- **Purpose:** used to assist in the assessment and diagnosis of personality disorders and clinical syndromes as outlined in the DSM-IV
- **Population:** 18 years old and older
- **Score:** Base Rate (BR) scores
 - 75 or above is considered a significant elevation, suggesting the presence of a personality trait or clinical syndrome
 - 85 or above indicates that the trait or syndrome is prominent in the individual's presentation
- **Time:** 175 true/false items; 25 to 30 minutes
- **Clinical Personality Scales:**
 - **(1) Schizoid:** aloof, apathetic, socially detached
 - **(2A) Avoidant:** socially inhibited, fear of rejection, feelings of inadequacy
 - **(2B) Depressive:** pessimistic, despondent, self-deprecating
 - **(3) Dependent:** submissive, clinging, needs reassurance
 - **(4A) Histrionic:** dramatic, attention-seeking, overly emotional
 - **(4B) Turbulent:** self-absorbed, arrogant, exploitative
 - **(5) Narcissistic:** grandiose, entitled, superficial charm
 - **(6A) Antisocial:** impulsive, irresponsible, disregard for rules, deceitful
 - **(6B) Sadistic:** hostile, unfeeling, enjoys dominating others
 - **(7) Compulsive:** perfectionistic, rigid, overly controlled
 - **(8A) Negativistic:** resentful,



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

rebellious, procrastinating

- **(8B) Masochistic:** self-sacrificing, self-deprecating, seeks painful experiences

- **Severe Personality Scales:**

- **(P) Paranoid:** suspicious, distrustful, guarded, believes other intended harm
- **(C) Borderline:** unstable mood, relationships, and self-image; impulsivity; fear of abandonment
- **(S) Schizotypal:** eccentric, peculiar thoughts and behaviors, social isolation

- **Clinical Syndrome Scales**

- (A) Generalized Anxiety
- (H) Bipolar Spectrum
- (D) Persistent Depression
- (B) Alcohol Dependence
- (T) Drug Dependence
- (H) Somatoform
- (R) Post-Traumatic Stress Disorder

- **Severe Clinical Syndrome Scales:**

- **(SS) Thought Disorder:** disorganized thinking, unusual perceptions
- **(CC) Major Depression:** severe depressive symptoms, anhedonia, hopelessness
- **(PP) Delusional Disorder:** fixed, false beliefs not amenable to reason

- **Validity Scales:**

- **Modifying Indices**

- **(X) Disclosure Scale** – indicating whether the respondent was overly guarded or overly open
- **(Y) Desirability Scale** – detects presenting oneself in an overly favorable or socially desirable light
- **(Z) Debasement Scale** – tendency to exaggerate problems or present oneself in an overly negative or pathological light

- **Random Response Indicators**

- **(V) Validity Scale** – contains

a number of improbable items which may indicate questionable result if endorsed

- **(W) Inconsistency Scale** – detects differences in responses to pairs of items that should be endorsed similarly

Millon Clinical Multiaxial Inventory 4th Edition (MCMI-IV)

- **Purpose:** to aid in the diagnosis and treatment planning for individuals in clinical settings

- Personality patterns and disorders aligned with the DSM-5
- Clinical syndromes, such as anxiety and depression
- Underlying personality dynamics that may be affecting a person's symptoms and behavior

- **Population:** 18 years and older who are undergoing psychological or psychiatric assessment or treatment.

- Its use is limited to individuals with a clinical referral question, and it is not recommended for use with the general population or for non-clinical purposes.

- **Score:** 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

- **Time:** 195 true/false items; 25 to 35 minutes

- **Clinical Scales:**

- **12 Clinical Personality Pattern Scales:** Schizoid, Avoidant, Melancholic,



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Sacuzzo, Miller & Lovler

Dependent, Histrionic, Turbulent,
Narcissistic, Antisocial, Sadistic,
Compulsive, Negativistic, and
Masochistic

- **3 Severe Personality Pathology Scales:** Schizotypal, Borderline, and Paranoid

- **10 Clinical Syndrome Scales:**

- **Clinical Syndromes (7 scales):** Generalized Anxiety, Somatic Symptom, Bipolar Spectrum, Persistent Depression, Alcohol Use, Drug Use, Posttraumatic Stress

- **Severe Clinical Syndromes (3 scales):** Schizophrenic Spectrum, Major Depression, Delusional

- **Validity Scales**

- **Invalidity (V)** – detects highly improbable responses that suggest an invalid protocol
- **Inconsistency (W)** – measures a person's tendency to answer similar item pairs inconsistently
- **Disclosure (X)** – assesses the level of openness or secretiveness in a person's responses
- **Desirability (Y)** – identifies a tendency to present oneself in an unrealistically positive light ("faking good")
- **Debasement (Z)** – measures an inclination to exaggerate problems or present oneself in an overly negative light ("faking bad")

Personality Assessment Inventory (PAI)

- **Purpose:** a comprehensive self-report instrument designed to provide information relevant to clinical diagnosis, treatment planning, and screening for psychopathology in adults
- **Population:** 18 years and older
- **Time:** 40 to 60 minutes
- **Domains Measured:**
 - **Clinical Scales**

- **(Som) Somatic Complaints** – measures preoccupation with physical health and somatization

- **(Anx) Anxiety** – measures general anxiety symptoms, tension, and worry

- **(Ard) Anxiety-Related Disorders** – assesses specific anxiety disorders

- **(Dep) Depression** – measures symptoms of depression, sadness, and low mood

- **(Man) Mania** – measures symptoms of elevated mood, grandiosity, and impulsivity

- **(Par) Paranoia** – suspiciousness, distrust, and beliefs of persecution

- **(Scz) Schizophrenia** – measures unusual thoughts, perceptions, and social alienation

- **(Bor) Borderline Features** – emotional lability, impulsivity, and relationship problems

- **(Ant) Antisocial Features** – antisocial behaviors, egocentricity, and lack of empathy

- **(Alc) Alcohol Problems** – identifies potential problems related to alcohol use

- **(Drg) Drug Problems** – identifies potential problems related to drug use

- **Treatment Consideration Scales**

- **(Agg) Aggression** – measures different forms of aggressive attitudes and behaviors

- **(Sui) Suicidal Ideation** – assesses thoughts and plans related to suicide, ranging from vague ideas to specific plans



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPLP

Main References: Cohen & Swerdlik, Kaplan & Sacuzzo, Miller & Lovler

- **(Non) Nonsupport** – reflects perceived lack of social support and isolation
- **(Str) Stress** – impact of recent and ongoing life stressors
- **(RxR) Treatment Rejection** – resistance or unwillingness to engage in psychological treatment
- **Interpersonal Scales**
 - **(Dom) Dominance** – assertive, controlling, or independent in relationships
 - **(Wrm) Warmth** – kind, empathetic, and affiliative in relationships
- **Validity Scales:**
 - **(ICN) Inconsistency** – random or careless responding
 - **(IFN) Infrequency** – identifies highly unusual or idiosyncratic responses, which can suggest exaggeration of symptoms or a misunderstanding of items
 - **(PIM) Positive Impression** – a respondent is presenting themselves in an overly favorable or virtuous light (“faking good”)
 - **(NIM) Negative Impression** – a respondent is exaggerating problems or presenting themselves in an overly negative light (“faking bad”)

- **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)

F. Attitude Test

Satisfaction with Life Scale

- **Purpose:** to assess satisfaction with people's lives as whole
- **Population:** 10 years and older
- **Time:** 1 minute
- **Domains Measured:** global life satisfaction

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

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



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

- **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

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

Scenotest [OUTDATED TEST]

- **Purpose:** to gain insight that the patient may consciously be withholding, or may not be able to express, due to emotional or physical abuse or for other complex reasons
- **Population:** all ages

- **Time:** untimed (approx. 30 minutes to complete)
- **Domains Measured:** understanding unconscious problems and personality structure in children

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



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

- 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
 - Year of Publication: 1965
- **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
- Year of Publication: 1967

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

A Detect Errors and Impacts in Test

Issues in Test Administration

- Procedures in Test Administration
 - Before the Test
 - Selecting tools that are appropriate to use
 - Secure tools (must not be known by the test taker in advance)
 - **Protocol** – the form or sheet or booklet which a test taker's responses are entered
 - During the Test Administration
 - Building up **rapport**—working relationship between examiner and examinee
 - After the Test Administration
 - Safeguarding the protocol
 - Writing report
 - Scoring and interpretation of the test result
- Examiner and the Subject
 - Different assessment procedures require different levels of training
 - Examiners should be aware that their rapport with test takers can influence the result
 - The test should be given in the language that the test taker feel is their best
 - The race should also be considered while administering the test



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

- Assessing People with Disabilities
 - Accommodation** – the adaptation of test, procedure, or situation or the substitution of one test for another, to make the assessment more suitable for an assessee with exceptional needs
 - Alternate Assessment** – an evaluative or diagnostic procedure or process that varies from the usual, customary, or standardized way of measurement

Issues in Intelligence Testing

- Flynn Effect** – is the progressive rise in the intelligence test score that is expected to occur on a normed intelligence test from the date when the test was first normed
 - Frog Pond Effect** – theory that individuals evaluate themselves as worse when in a group of high-performing individuals
- Culture Bias of Testing**
 - Bias in Testing** – presence of systematic error in the measurement of certain factors
 - Culture Free** – attempt to eliminate culture so nature can be isolated without cultural influences and biases
 - Culture Fair** – minimize the influence of culture with regard to various aspects of the evaluation procedures
 - Culture Loading** – the extent to which test incorporates the vocabulary, concepts, traditions, knowledge, etc. of particular time

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
- 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*
 - $X = T \text{ (true score)} + E \text{ (error)}$
- 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.



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

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



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

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

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

		Does The Effect Exist?	
		Effect Exists	Effect Doesn't Exist
Was The Effect Observed?	Effect Observed	Hit True Positive	False Alarm False Positive Type I Error
	Effect Not Observed	Miss False Negative Type II Error	Correct Rejection True Negative

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

- **Reactivity** – when evaluated, the behaviors increase
 - Hawthorne Effect
- **Drift** – moving away from what one has learned going to idiosyncratic definitions of

behavior

- Subjects should be retained in a point of time
- **Contrast Effect** – cognitive bias that distorts our perception of something when we compare it to something else, by enhancing the difference, decrease one's performance
- **Expectancies** – tendency for results to be influenced by what test administration expect to find
 - **Rosenthal/Pygmalion Effect** – test administrator's expected results influences the result of the test
 - **Golem Effect** – negative expectations decreases one's performance
 - **Galatea Effect** – self-expectations result in higher level of performance
- **Rating Errors** – intentional or unintentional misuse of scale regardless of the performance of the test taker
 - **Leniency Error** – the rater is lenient in scoring (generosity error)
 - **Severity Error** – rater is too strict in scoring
 - **Central Tendency Error** – rater's rating would tend to cluster in the middle of the rating scale
 - **Halo Effect** – tendency to give high score due to failure to discriminate among conceptually distinct and potentially independent aspects of a ratee's behavior
 - **Horn Effect** – opposite of halo effect
 - To overcome the rating error is to use rankings
- **Fundamental Attribution Error** – tendency to explain someone's behavior based on internal factors such as personality or disposition, and too underestimate the influence the external factors have on another person's behavior,



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

blaming it on the situation

- **Barnum Effect** – people tend to accept vague personality descriptions as accurate descriptions of themselves (Aunt Fanny Effect)
- **Bias** – factor inherent in a test that systematically prevents accurate, impartial measurement
 - Prejudice, preferential treatment
 - Prevention during test development through a procedure called *Estimated True Score Transformation*
- **Social Desirability** – tendency to say good things about yourself or to mark items that you believe will be approved by the examiner
 - Impression Management
 - Faking Good/Bad

D

Levels of Interpretation

Level I – there is a minimal amount of any sort of interpretation

- Minimal concern with intervening processes
- Data are primarily treated in a sampling or correlate way
- No concern in large-scale selection testing
- For psychometric approaches

Level II

- Descriptive Generalization
- **Hypothetical Construct** – the assumption of an inner state which goes logically beyond the description of visible behavior

Level III – the effort to develop a coherent and inclusive theory of the individual life or a “working image” of the patient

- The clinician attempts full scale exploration of the individual personality, psychosocial situation, and developmental history

Extra Test Behavior

- Observations made by an examiner regarding what the examinee does and how the examinee reacts during the course 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



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

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



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

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
- 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



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Sacuzzo, Miller & Lovler

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



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

- 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

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



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

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 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



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

- 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



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Sacuzzo, Miller & Lovler

- **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
- **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



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

→ 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

HEXACO Model of Personality Traits – is a six-dimensional model of personality structure that was developed by psychologists Michael C. Ashton and Kibeom Lee

- It is an alternative to the widely known Big Five model of personality and is based on extensive research using a "lexical approach," which involves analyzing the words people use to describe themselves and others



6 Broad Personality Traits

→ (H) Honesty-Humility – individuals who score high on this trait are sincere, fair, and modest; they are less likely to manipulate others for personal gain, avoid breaking rules, and are not interested in lavish wealth or social status

- ◆ Conversely, people with low scores are prone to deceit, greed, and a sense of entitlement
- ◆ Facets: Sincerity, Fairness, Greed-Avoidance, Modesty

→ (E) Emotionality – high emotionality is characterized by a tendency to experience anxiety, fear, and a need for emotional support. Individuals with high scores may also be sentimental and empathetic

- ◆ People with low scores are more emotionally stable, brave, and less prone to worry or anxiety
- ◆ Facets: Fearfulness, Anxiety, Dependence, Sentimentality

→ (X) Extroversion – describes a person's tendency to be sociable, lively, and outgoing; high scorers are socially confident, cheerful, and enjoy being the center of attention.

- ◆ Low scorers are more reserved, withdrawn, and prefer solitude
- ◆ Facets: Social Self-Esteem, Social Boldness, Sociability, Liveliness

→ (A) Agreeableness – people who score high on this trait are patient, tolerant, and gentle; they are willing to compromise and forgive

- ◆ Low scorers are more likely to be



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPLP

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

critical, stubborn, and quick to anger

- ◆ Facets: Forgivingness, Gentleness, Flexibility, Patience

→ **(C) Conscientiousness** – reflects a person's level of organization, diligence, and discipline; high scorers are meticulous, thorough, and hard-working

- ◆ Low scorers are more disorganized, careless, and impulsive
- ◆ Facets: Organization, Diligence, Perfectionism, Prudence

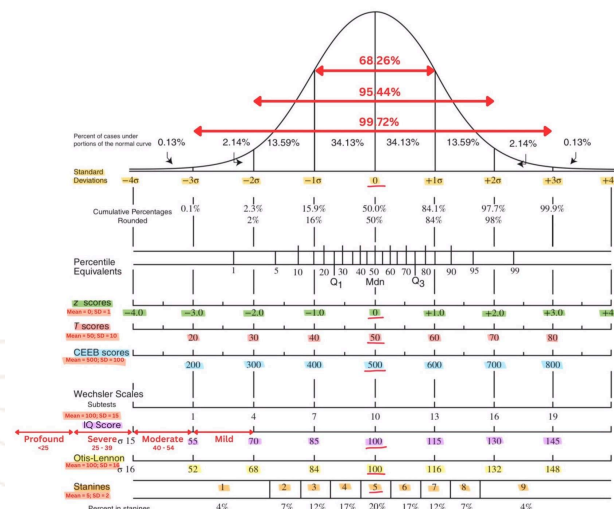
→ **(O) Openness to Experience** – describes a person's intellectual curiosity and appreciation for art, nature, and new ideas; high scorers are imaginative, creative, and unconventional

- ◆ Low scorers are more conventional, practical, and less interested in artistic or intellectual pursuits
- ◆ Facets: Aesthetic Appreciation, Inquisitiveness, Creativity, Unconventionality

HEXACO vs. OCEAN

Feature	HEXACO Model	OCEAN (Big Five) Model
Number of Factors	Six	Five
Additional Factor	Honesty-Humility: This is the most significant addition. It describes a person's tendency to be sincere, fair, and modest. People low on this trait are prone to manipulation, greed, and a sense of entitlement.	None
Emotionality vs. Neuroticism	Emotionality (E): This factor emphasizes fear, anxiety, and the need for emotional support.	Neuroticism (N): This factor focuses more broadly on emotional instability, including anger, depression, and moodiness, in addition to anxiety.
Agreeableness Content	Agreeableness (A): The content of this factor is slightly different. The HEXACO version is more about patience and tolerance, with a focus on a person's anger and willingness to forgive.	Agreeableness (A): The Big Five version includes more of the traits related to altruism, trust, and sympathy, while some of the facets related to anger and manipulation are found in Neuroticism and other factors.

NORMAL CURVE (got this from rpmtwt)



STATISTICAL TESTS

COMPARING TWO MEAN GROUPS (t-TESTS)

- **Z-Test** – compares a sample mean to a population mean or compares two means
 - Used when the population standard deviation is known and the sample size is large, typically more than 30 samples
- **Student's t-Test** – compares the means of two independent groups (e.g., Treatment A vs. Treatment B)
 - Assumes that the two groups have equal variances (homogeneity of variance)
- **Paired t-Test** – compares the means of two related samples (e.g., the same individuals before and after the treatment)
 - Focuses on the mean of the differences between the pairs
- **Welch's t-Test** – compares the means of two independent groups
 - Does not assume equal variances (more robust than the Student's t-Test when variances are unequal)

COMPARING ≥ 3 MEAN GROUPS (P & NP)

- **ANOVA (Analysis of Variance)** – determines if there is a statistically significant difference between the means of three or more independent groups. The null hypothesis states all mean groups are equal



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

- **One-Way ANOVA** – determine if there are significant differences between the means of the groups created by the single factor
 - One factor with three or more levels (groups)
 - Independent groups (between-subjects)
 - **Two-Way ANOVA** – examine the main effect of each factor and the **interaction effect** between the two factors on the dependent variable
 - Two factors, each with two or more levels.
 - Independent groups (between-subjects)
 - **Repeated Measures ANOVA** – determine if there are significant differences in means over time or under different conditions when the same people are used
 - One or more factors (often time or condition)
 - Related groups (within-subjects), where the same subjects are measured multiple times
 - **Factorial ANOVA** – general term for any ANOVA with two or more factors (e.g., a two-way ANOVA is a type of factorial ANOVA)
 - Two or more factors
 - Independent groups (between-subjects)
 - **ANCOVA (Analysis of Covariance)** – an extension of ANOVA that includes one or more covariates (continuous variables that are not the main variable of interest but might influence the dependent variable)
 - To statistically control for the effect of a continuous variable (the covariate) on the dependent variable, thereby increasing the power and precision of the test on the main factor(s)
 - **MANOVA (Multivariate Analysis of Variance)** – an extension of ANOVA used when there are two or more continuous dependent variables
 - To determine if there are differences between independent groups on a set of dependent variables simultaneously
 - This controls the family-wise error rate (the risk of making a Type I error) that would increase if multiple separate ANOVAs were conducted
 - **MANCOVA (Multivariate Analysis of Covariance)** – combination of MANOVA and ANCOVA
 - It includes two or more continuous dependent variables and one or more covariates
 - To compare groups on multiple dependent variables while controlling for the effect of one or more continuous covariates
 - **Kruskal-Wallis Test** – determines if there is a statistically significant difference between the medians (or distributions) of three or more independent groups
 - This is the alternative to ANOVA when the data is non-normal or ordinal
- ## TESTS FOR CATEGORICAL DATA (FREQUENCIES)
- **Chi-Squared (χ^2) Test** – the most common test for categorical data. It assesses the association (independence) between two categorical variables (e.g., is gender independent of political party preference?)
 - **Fisher's Exact Test** – used specifically for 2×2 contingency tables when the sample size is small, and the χ^2 test's assumption of large expected cell counts is violated. It computes the exact probability
- ## NONPARAMETRIC ALTERNATIVES TO t-TESTS
- **Mann-Whitney U Test** (Independent Samples



PSYCHOLOGICAL ASSESSMENT

BLEPP/PPL

Main References: Cohen & Swerdlik, Kaplan & Saccuzzo, Miller & Lovler

t-Test) – compares the distributions (often interpreted as medians) of two independent, non-normally distributed groups

- **Kruskal-Wallis Test** (One-way ANOVA) – compares the distributions (often interpreted as medians) of three or more independent, non-normally distributed groups
- **Spearman's Rank Correlation (ρ)** (Pearson's Correlation) – measures the strength and direction of a monotonic relationship between two variables, using the ranks of the data instead of the raw value

REGRESSION AND CORRELATION (MODELING RELATIONSHIPS)

- **Pearson's Correlation (r)** – measures the linear association between two variables. The result is a number between -1 (perfect negative relationship) and +1 (perfect positive relationship)
 - Both variables must be continuous
- **Simple Linear Regression** – models the linear relationship to predict a continuous outcome (Y) from a single predictor (X)
 - Dependent Variable (Y): Continuous.
 - Independent Variable (X): Continuous or Categorical
- **Multiple Regression** – models the linear relationship to predict a continuous outcome (Y) from two or more predictors ($X_1, X_2, \dots$)
 - Dependent Variable (Y): Continuous.
 - Independent Variables (X): Any type
- **Logistic Regression** – predicts the probability of a categorical outcome (e.g., 0 or 1, Yes or No) occurring based on one or more predictors
 - It uses the log-odds (logit) transformation
 - Dependent Variable (Y): Categorical/Binary. Independent Variables (X): Any type

END