Historical Perspectives

Early Testing *Greeks* 2500 years ago *Chinese* 2000 years ago *Francis Galton* credited with launching the testing movement (e.g. hereditary, statistics) *Wilhelm Wundt* credited with founding the science of psychology. Interested in identifying factors of intelligence that were common/innate to all human beings *James McKeen Cattell* expanded testing to include memory/other simple mental processes

1900-1920 *Binet*: first person to consider Intelligence Quotient (IQ) *Binet-Simon (1905)* scale focused on assessing judgment, comprehension & reasoning

Ratio of mental age to chronological age (IQ). Revised test became known as Stanford-Binet scale (1916) *World War I* – group testing (Army Alpha & Army Beta)

Frank Parsons "father of guidance". Career counseling: 1) understand client 2) understand world of work 3) match client to appropriate occupation in the world of work

1920s and 30s Aptitude tests developed for finding industrial personnel. Development of vocational counseling instruments. *Stanford Achievement Test (1923)* first standardized achievement battery. First edition of Mental Measurements Yearbook (1939)

1940s and 50s Dissatisfaction with existing personality instruments. Projective techniques became popular (Rorschach). MMPI developed (early 1940)

Standardized achievement tests well established in public schools. Criticisms of assessments began to emerge. Need for standards (APA). Need for centralized test publication, electronic scoring

1960s and 70s Widespread public concern of assessment (standardization etc.) **70s**: Grassroots movement for minimum competency testing in high school graduates

Current thoughts on mental Disorders: Associated with significant distress and limitations in social, occupational, or other important aspects of life, Refers to significant disturbances in cognition, emotions, and/or behaviors, Mental disorders are classified in the Diagnostic and Statistical Manual of Mental Disorders

Terms & Tools

Assessment involves some type of measurement and gathering samples of behavior in an objective, standardized, and systematic manner. It is a constant process.

Assessment: A procedure for gathering client information that is used to facilitate clinical decisions, provide clients w/info or evaluation purposes

Appraisal: Another term for assessment **Testing**: Another term for intelligence test **Instruments**: Test, scales, checklists, & inventories

The correct order of the following scales of measurement, from least precise to most precise *Nominal, ordinal, interval, ratio*

Treatment monitoring systems detect potential treatment failure

Qualitative data: uses open-ended questions or interviews

Quantitate data: uses evaluating paperwork, assessments, etc.

Assessment Tools

Individual vs. group, Objective vs. subjective, Verbal vs. nonverbal. Performance tests require the manipulation of objects with minimal verbal influence (e.g. doing a puzzle). Language and culture influences assessment,

Speed vs. power *Power*: items in the experiment very in difficulty. *Speed*: examines the number of items completed in a specified time period.,

Terms & Tools (cont)

Cognitive vs affective *Cognitive instruments* assess cognition: perceiving, processing, abstract and concrete thinking, and remembering. *Includes*: Intelligence, Achievement, Aptitude.*Affective instruments* assess interest, attitudes, values, motives, temperaments, and the noncognitive aspects of personality. MMP-2. *Projective techniques*: individuals respond to ambiguous stimuli like inkblots. Like many other tests, when administering the Rorschach the administrator needs to record verbatim the content of the responses

Statistics

Measurement scales: to measure some aspect of the client, methods depend on type of scale *Nominal*: classifying by name based on characteristics. We get a count of percentage (i.e. political parties, male:female amount in class); categories; nominal=naming categories *Ordinal*: a measure of magnitude ranks; smaller/larger (i.e. ranking behaviors from best to worse in a classroom); gives you an idea of what goes first, what goes second; ordinal=order *Interval*: ranking units have equal intervals but not true zero (i.e. intelligence scores) *Ratio*: has interval but with meaningful zero (i.e miles per hour, weight)

Measures of central tendency: *Mode*: most frequent score (bimodal, multimodal) *Median*: evenly divides scores into two halves *Mean*: arithmetic average of the scores

Measures of variability *Range*: highest score minus lowest score *Variance*: amount of spread in a group of scores. The greater the spread, the greater the variance *Standard Deviation*: square root of variance

Indications of variability how far apart data points lie from each other and from the center of a distribution

Types of Scores in Normal Distributions 1.*Raw Scores* 2. *Percentile scores/Percentile ranks*: position relative to standardized sample 3.*Standard score*s: the common language. Converts raw scores, has a set mean and standard deviation. Express how far a score is from the mean. **Examples**: *z scores* (mean is always 0 and standard deviation is always 1), *T scores* (always has a mean of 50 and a SD of 10), *Stanines* (rarely used), *Age/grade-equivalent scores* (highly criticized unless used, with caution, with kids, especially in a school setting).

Shapes and Types of Distributions *Skewed Left* the bulk of the data values (including the median) lie to the right of the mean, and there is a long tail on the left side *Skewed Right* the bulk of the data values (including the median) lie to the left of the mean, and there is a long tail on the right side *Symmetric/Unimodal* approximately half of the data values lie to the left of the mean, and approximately half of the data values lie to the right of the mean *Uniform/rectangular* All entries or classes in the distribution have equal or approximately equal frequencies *Bimodal* a bimodal distribution has two modes, or two distinct clusters of data *Multimodal* more than one mode *Bell Curve* tend to have a central, normal values, as peak with low and high extremes tapering off relatively symmetrically on either side

Shapes and Types of Correlations *Correlation* describes the relationship between variables. It can be described as either strong or weak, and as either positive or negative *Positive Linear Correlation* the variable on the x -axis increases as the variable on the y -axis increases *Negative Linear Correlation* when one variable increases as the other variable decreases *Non-linear Correlation (known as curvilinear correlation)* a relationship between variables but the relationship is not linear/straight *No Correlation* when there is no pattern that can be detected between the variables

Reliability and Validity

tests can be: reliable not valid, low validity and low reliability, not reliable and not valid, and both reliable and valid

Reliability

Reliability and Validity (cont)

A test is reliable to the extent that whatever it measures, it measures it consistently. three major categories of reliability for most instruments: *test-retest, equivalent form, and internal consistency*

a minimum reliability of .70 is required for attitude instruments. Many tests, such as achievement tests, strive for .90 or higher reliabilities.

Test-retest (stability: measures error because of changes over time) The same instrument is given twice to the same group of people. The reliability is the correlation between the scores on the two instruments.

Equivalent-Form (Parallel or Alternate-Form) Method (measures error because of differences in test forms) Two different versions of the instrument are created. The scores on the two instruments are correlated to calculate the consistency between the two forms of the instrument.

Internal-Consistency Method (measures error because of idiosyncrasies of the test items) *Split-Half* A total score for the odd number questions is correlated with a total score for the even number questions (or first half & second half) with the Spearman-Brown prophecy formula, Kuder--Richardson Formula, or Cronbach's Alpha

Scoring Agreement (measures error because of the scorer) *Interrater Reliability* two judges can evaluate a group of student products and the correlation between their ratings can be calculated *Percentage Agreement* two judges can evaluate a group of products and a percentage for the number of times they agree is calculated

Inconsistencies test taker issues, test itself, testing conditions, test scoring

Validity

Content validity indicates the extent to which items adequately measure or represent the content of the property or trait that the researcher wishes to measure.

Construct validity indicates the extent to which a measurement method accurately represents a construct (e.g., a latent variable or phenomena that can't be measured directly, such as a person's attitude or belief) and produces an observation, distinct from that which is produced by a measure of another construct. Common methods to assess construct validity include, but are not limited to, factor analysis, correlation tests, and item response theory models

Criterion-related validity indicates the extent to which the instrument's scores correlate with an external criterion (i.e., usually another measurement from a different instrument) either at present (concurrent validity) or in the future (predictive validity). A common measurement of this type of validity is the correlation coefficient between two measures.

Standardized Scores

A SD is a measure of how dispersed the data is in relation to the mean. Low standard deviation means data are clustered around the mean, and high standard deviation indicates data are more spread out.

Standardized and Non-standardized Testing

A standardized test is a test that is administered and scored in a consistent, or "standard", manner.

Types of tests: Cognitive, Personality (structured or non-structured), Neuropsychological, Achievement/School

Standardized and Non-standardized Testing (cont)

Projective tests (aka unstructured): Thought to uncover more of client's unconscious and thus providing an indication or covert or latent traits (more difficult to "fake" responses). Extensive training is needed to use them appropriately. They are more subjective to interpretations. Strengths: more difficult to fake, can sometimes identify more complex themes and multidimensional aspects of personality, helpful with children and nonverbal client, can serve as an effective method of establishing rapport. Limitations: lower reliability evidence, more caution needed when interpreting results, meager validation information, lack of normative data, can be dangerous with untrained users.

Construction (of stories/TAT) Thematic Apperception Test (TAT) / Child's Apperception Test (CAT) Projective technique consisting of a series of pictures. Examinee needs to create a story about what they believe is occurring in the situations or events depicted in pictures. Instructions are to create a story with a beginning, middle, and end. Want client to describe what the characters are feeling. There is no "abnormal" response – account for gender, sexual identity, etc., The client can exercise control over the story and choose to include or emphasize particular contents and /or distort and omit other contents. No simple conclusions can be made as interpretation is complex.

Completion (of sentences, cartoons/sentence completion, Rotter Incomplete Sentences Blank) Incomplete sentences test: a test that provides a series of sentence beginnings that the client is to complete which can be interpreted qualitatively for both quality of expression and vocabulary, as well as for the psychological content and themes

Association techniques (Rorschach/inkblot, word associations)*Arrangement/selection* (toys, verbal options, puppets, sand play) *Expression* (draw a person, house/tree/person, kinetic family test)

Objective tests (aka structured tests) give you options to choose from *test types*: Beck, MMPI, Self-report measures

Cognitive, Neuropsychological, Achievement/School are all related (measure some form of intelligence or functioning). Different tests measure different areas of the brain. One typically starts with cognitive testing, then refers to neuropsychological testing as needed

Beck: produce descriptive answers that do not need interpretation. Used for: Detecting the possible presence of depression. Determine the baseline level of depression severity for a client. If repeated, could be used to determine the effectiveness of interventions targeted to treat the depression.

Norm-Referenced and Criterion-Referenced Instruments *Norm-referenced instrument*: individual's score is compared to performance of others who have taken the same instrument (norming group). Example: personality inventory. Evaluating the norming group (size, sampling, representation) *Criterion-Referenced Instrument*: individual's performance is compared to specific criterion or standard. Example: third-grade spelling test. How standards are determined: common practice, professional organizations or experts, empirically determined

Referrals

Referrals should have an electronic/written trail. Clients being referred need to understand why they are being referred out.

If a family is unwilling to be seen as a whole, refer out (unless the identified patient or family member is in danger – in which case, seek individual therapy. Especially important if you sense abuse).

Tests

Academic/Educational

Tests (cont)

Learning and environment influence our cognitive processes . In theory, intelligence is stable. Infants and young children have the least stable intelligence test scores. Intelligence scores test to be related to academic performance. environmental factors may hinder intellectual development/performance: socioeconomic, language/culture, trauma/stress, parent involvement, sleep, etc.

Survey Battery Tests: Tests, usually given in school settings, which measure broad content areas and often used to assess progress in school ex: NAEP, Stanford 10, ITBS, Metropolitan Achievement Test

Diagnostic Tests: Tests that assess problem areas of learning and often used to assess learning disabilities ex: PL94-142, IDEA, WRAT4, WIAT-III, PIAT, etc.

Readiness Tests: Tests that measure one's readiness for moving ahead in school and often used to assess readiness to enter first grade ex: MRT6, Gesell Development Observation

Cognitive Ability Tests: Tests that measure a broad range of cognitive ability. These tests are usually based on what one has learned in school and are useful in making predictions about the future ex: OLSAT 8, CogAT, ACT, SAT

Wechsler scales: *WAIS is the most widely used intellectual assessment for adults*; Administration time is 60-90 minutes for core subtests; Normative population: US, ages 16-90; Norms have been developed for special groups (e.g. mild cognitive impairment, borderline intellectual functioning, TBI, ADHD, AD, reading disorder, autism, depression. Includes adult, child and preschool tests

Stanford-Binet intelligence Scale, Kaufman Instruments are similar to WAIS	Woodcock-Johnson II Complete Battery: has cognitive and achievement, and now also has oral language

Additional Individual Assessments: Peabody picture, test of nonverbal intelligence (TONI), differential ability scales, Slossen intelligence test, Raven's standard progressive matrices

Career

Domains: personality tests, interest inventories, skills assessments, and value assessment	MBTI, SII, Holland

Personal and Social Development

Informal Assessments: Observation, Interviewing **Formal Assessments**: MBTI, MMPI-2, NEO-PI-3	**Projective Techniques** *Categories*: Associations (Rorschach), Construction (Thematic TAT), Completions (Rotter Incomplete Sentences), Arrangement/selection (Sand, Puppets, Toys), Expression (Drawing)

MMPI-2 Hypochondriasis, Depression, Conversion Hysteria, Psychopathic Deviate, Masculinity-Femininity, Paranoia, Psychasthenia, Schizophrenia, Hypomania, Social Introversion

NEO-PI-3 Surgency (or Extroversion), Agreeableness, Conscientiousness, Emotional Stability or Neuroticism, Intellect or Openness to Experience

MBTI Extroversion – Introversion, Sensing – Intuition, Thinking – Feeling, Judging – Perceiving

Other

A developmental history: Assists in understanding the physical, psychological, interpersonal, innate disposition of parent and child, Assists in understanding the developmental growth of a child and parents' expectations, Allows to assess functioning overtime (temporary vs. temperamental)

Norm-referenced instruments are MOST useful for comparing an individual's performance with other individuals

Brief instruments are self report measures that reflect clients' own perception of difficulties and take little administration time and are useful for treatment planning

Tests (cont)

The Exner Comprehensive System, which examines location, determinants, and content, is associated with the Rorschach Inkblot Test

Ethical Issues & Multicultural Consideration

Consider which assessment method would be best suited to client.

Consider professional limitations & which instruments counselor can ethically administer and interpret. Reliability and validity, fairness, user qualification, practical issues, interpretation and scoring materials

Communication of results is most important. It requires specialized knowledge and competencies. Considerations include:1. Use visual aids to explain technical terms, 2. Use descriptive terms rather than numerical scores 3. Tentative interpretations rather than absolutes 4.Discuss limitations of assessment in non-technical terms 5.Monitor client reactions during interpretation

Multicultural Considerations

Language and culture influences as assessment (e.g. verbal vs. nonverbal assessment tools) Examine characteristics of norming group (e.g. race, ethnicity, disability, gender) For referrals- if there If there are unique considerations in a culture, include small print

First Meeting

1) Assess the client problem(s). Find out highest priority in terms of client's problem from their own words/view. Explore each significant problem from multiple perspectives; gather specific information; assess each problem's intensity; assess degree to which client believes each problem is changeable; identify methods the client has previously used to solve the problem.

2) Conceptualize & define client problem(s) **3)** Select & implement effective treatment(s), **4)** Evaluate counseling

Interviewing skills/techniques: open questioning; paraphrasing; clarifying; reflecting; interpreting; summarizing. **Motivational interviewing**: express empathy

Mental Status Examination: used to describe client's level of functioning and self-presentation. Generally conducted during initial session/intake interview. **Appearance, attitude, and activity; mood and affect; speech and language; thought process, thought content and perception; cognition; insight and judgement**. Stay away from adjectives – they are our interpretations. MSE should be about the FACTS. Provider's observations at the time of the encounter. *Helps understand which parts of the individual's presentation is related to stress (external) vs personality (characterological)*

Assessment is integral to counseling: Helps share info with collaterals, Understanding guides intervention, Some patients uncomfortable with verbalizing, helps counselor gain more info into underlying issues, Insurance, Having pre-, mid-, and post-assessment to measure outcomes

Trauma and Abuse

Goal is to assess the POSSIBILITY of some type of abuse and not to determine if abuse has occurred. Areas of assessment (these should be patterns… not necessarily one time occurrences): *Physical* (characteristics of the child), *Behavioral Signs* (e.g. sleepy, stealing food, hungry, not attending school) , *Caretaker characteristics* (e.g. chaotic)

If abuse is disclosed, you should gather information, consult and decide based on nature of incident, age, and other factors

Columbia Suicide Severity Rating Scale assesses risk of aggression or danger to others, self-inflicted harm and/or suicidality

Environmental & Behavioral Observations

The manner of gathering behavior samples should be objective, standardized, and systematic *Subject's behaviors*: Verbal remarks; nonverbal reactions *Record behavioral observations*: long pauses, blushes, exclamations, examiner can ask clarifying questions

Environmental & Behavioral Observations (cont)

Formal Behavioral Assessments: *Behavior Assessment System for Children, Second Edition (BASC-2)* Adaptive and maladaptive behaviors; Multimethod and multidimensional assessment; Five versions; Teacher Rating Scale (TRS); Parent Rating Scale (PRS); Student Self-Report of Personality (SRP) Structured Developmental History (SDH); Student Observation System *Achenbach System of Empirically Based Assessment* Used with ages 1.5 to 90; Most common is Child Behavior Checklist System/6-18; 4 page assessment; Boys and girls scored separately; Adaptive functioning and strength scales, syndrome scales, and DSM-5-oriented scales

Tests for Diagnosis

Symptoms Checklist: any self-report inventory used to assess the severity of symptoms of a given disorder, as in the Beck Depression Inventory, or across a range of disorders, as in the Symptom Checklist-90–Revised.

Symptom Checklist-90 Revised: nine descriptive scales: somatization, obsessive-compulsive, interpersonal, depression, anxiety, hostility, phobic, paranoid, psychoticism

Ways to Diagnose Include: Symptoms checklist, DSM Information and assessment recommendations, the clinical interview, personality tests, intelligence tests, neuropsychological tests, physical & lab testing

All these are examples of reasons why patients are referred to neuropsychological testing: To assess if a child who is struggling academically in school has a learning disability, To provide recommendations regarding the capacity to return to work or to receive vocational training, To make recommendations regarding living situation for patients with cognitive, motor, and /or behavioral deficits

The State-Trait Anxiety Inventory (STA-I) measures both transitory anxiety and more stable personality features that predispose a client to more chronic levels of anxiety

Working with Couples

When interviewing/assessing couples it is useful to know how they negotiate arguments and to encourage "I " statements

In assessing couples, the focus in typically on the relationship and couple's distress

Special Topics

Suicide

Most persons who really want to end their lives, will find a way to do so.

A client has expressed some suicidal ideations and you want to explore further. The best question: When you think about hurting yourself, how do you imagine doing it?

Substance Use

In assessing potential substance abuse, we ask ourselves: Why is the person using, What happens when the person uses, When is the person using

Harm by a client

These are ways to minimize getting hurt by a client: Knowing security procedures at my agency, Not blocking the exit in the office/ place of meeting, Keeping some distance if a client is of concern

Other

Substantial research suggests that minority students tend to be overrepresented in special education

Clients who receive interpretation of test results tend to have a better counseling outcome than those who do not receive an interpretation

Special Topics (cont)

A counselor who is assessing a client not represented in the instrument's norming group should carefully consider how the information is going to be used and interpret the results very cautiously	A mark of a good psychological report is that it involves an overview of the client and a comprehensive interpretation of the assessment results
The first step in identifying appropriate assessment strategies is to identify what information is needed	With respect to high-stakes testing, both the American Psychological Association (APA) and the American Education Research Association (AERA) contend that tests should not be used as a single measure of student performance, or basis for high-stakes decisions

Rorschach, TAT, and Wechsler Intelligence test need to be administered,scored, and interpret by a trained psychologist