Self Report Inventories: Beck Depression Inventory

• 0  I do not feel sad.
• 1  I feel sad.
• 2  I am sad all the time and can’t snap out of it.
• 3  I am so sad or unhappy that I can’t stand it.
Self Report Inventories: Beck Depression Inventory

- 0 I don’t feel I am worse than anybody else.
- 1 I am critical of myself for my weaknesses or mistakes.
- 2 I blame myself all the time for my faults.
- 3 I blame myself for everything bad that happens.
Self Report Inventories: Beck Depression Inventory

• 0 I don’t have any thoughts of killing myself.
• 1 I have thoughts of killing myself, but I would not carry them out.
• 2 I would like to kill myself.
• 3 I would kill myself if I had the chance.
Limitations of Self-Report Inventories

• Response bias
Projective Techniques: Rorschach Inkblot Method
Projective Techniques:
Rorschach Inkblot Method
Limitations of Projective Techniques

• Reliability of scoring
• Validity of interpretations
Behavioral Observations: Peer Interaction Task (Dishion et al.)

- Plan an activity together (e.g., party)
- Solve 4 problems that have occurred within the past month:
  - A problem for the study boy/girl related to getting along with parents
  - A problem for the study boy/girl related to getting along with peers
  - A problem for the friend related to getting along with parents
  - A problem for the friend related to getting along with peers
Behavioral Observations: Peer Interaction Task (Dishion et al.)

• Coding
  – Rule-breaking (verbal behavior and gestures that contain some violation of legal and/or conventional norms of conduct)
  – Normative (anything not coded as rule-breaking)
Limitations of Observational Methods

- Sample of behavior
- Contrived situation
Psychophysiological Measures: Sexual Arousal Among Sexual Offenders (Hall, 1989)
Sexual Arousal Among Sexual Offenders (Hall et al., 1988)

The bar chart illustrates the level of sexual arousal among different scenarios:
- **Consenting adult**
- **Consenting child**
- **Child rape**
- **Child assault**

The chart shows the comparative levels of arousal (in millimeters) across these scenarios.
Sexual Arousal Among Normal Men (Hall, 1995)
Limitations of Psychobiological Measures

• Physiological responses do not necessarily correspond to psychological states
Unobtrusiveness and Reactivity of Measures

Anonymity may decrease response bias.

Filler or buffer items to disguise the purpose of the test.

Computerized assessment may be less reactive.
Unobtrusive measures

• Techniques
  – Observation in naturalistic settings
    • How is observation accomplished?
    • Frequency of target behavior
    • Lack of standardization
  – Observation in contrived situations - Stimulus is presented
    • Participant suspiciousness
  – Archival records (e.g., criminal records)
    • Selectivity in what is recorded
  – Physical traces (e.g., library vandalism)

• Ethical issues
  – Informed consent
Use of Multiple Measures and Modalities

• Multiple measures are usually needed to assess a construct
  – Most constructs have multiple components (e.g., affective, cognitive, physiological)
  – Are there instances where a single measure captures a construct?
Multiple Measures

- Consistencies, inconsistencies
- Synchrony and desynchrony of changes in measures

![Bar graph comparing outpatient and institutional measures]
Multitrait Multimethod Matrix

- **Method variance**
  - Correlations will be high within one method (e.g., self-report, behavioral observations)
  - Correlations of the same trait across methods support the validity of the measurement (concurrent validity)
- **Trait variance**
  - Different constructs should not be correlated (discriminant validity)
  - Correlation may suggest shared trait variance
Method Variance

- **Anger**
  - Variance
  - Concurrent Validity
  - Self-report
  - Other report
  - Observation
Assessment Research with Ethnic Minorities (Okazaki & Sue)

- Definition of ethnicity
  - Underlying psychological variables
- Individual vs. group differences
- Sampling
  - N, representativeness
  - Within-group heterogeneity
- Equivalence of measures
Assessment and Evaluation of Interventions
Assessing the Clinical Significance of Changes

• Practical value
• Most changes are a matter of degree rather than complete elimination

• Clinical significance
  – No longer meets diagnostic criteria
  – 2 SDs from the mean of the dysfunctional sample
Clinical Significance: Nonoverlapping distributions

Frequency

Control Group  Treatment Group
Comparison Methods

• Normative
  – Who is the normative group?
  – Are peer norms appropriate?
  – How do others view the problem?
  – Social impact measures
    • e.g., school referrals, arrest rates, hospitalizations
Comparison Methods

• Ipsative
  – Does impairment still exist?
  – Is the change clinically significant?
  – Does the person feel better?
  • A small change may make a big difference
Value of Treatment

- Disseminability
  - Extension to new settings
- Cost
  - Comprehensiveness
  - Cost of treatment vs. no treatment
- Acceptability
Assessment During the Course of Treatment

• Pre-post assessment most common

• Assessment throughout the course of treatment
  – May reveal the mechanisms of change (e.g., therapeutic alliance)
  – May reveal when changes occur

• Follow-up assessment
  – What are the lasting effects of treatment?
  – Attrition may prevent accurate follow-up assessment
  – Ongoing contact with participants may prevent attrition
Statistical Methods of Data Evaluation
Statistical Significance

• $p < .05 = 5$ times by chance out of 100
• Statistical significance is a function of sample size
  – Large samples make trivial differences significant
• Statistical significance does not address the strength or importance of an effect
Alternatives or supplements to significance tests

- Strength of effect – $r$, $r^2$, $R$, $R^2$
  - Small effect size, $r = .1$; medium effect size, $r = .3$; large effect size, $r = .5$
  - Effect size is not equivalent to importance

- Confidence intervals – likelihood that the effect size falls within a particular range
  - For $p = .05$, 95% CI = Mean $\pm$ 2.58 x SEM

- Meta-analyses
  - Effect sizes of multiple studies are combined
  - Studies need similar measures for inclusion
  - Quality of studies may vary
  - Individual studies may be weighted (e.g., by $N$, quality)
  - Moderators can be analyzed (e.g., groups for which effects are strongest)
Meta-analysis of 12 sex offender treatment studies (Hall, 1995)
Moderator of Treatment Effects (Hall, 1995)

The bar graph compares outpatient and institutional treatment effects. The x-axis represents the settings: Outpatient on the left and Institutional on the right. The y-axis shows the correlation coefficient (r) ranging from 0 to 0.3. The bar for Outpatient reaches up to 0.3, indicating a higher correlation coefficient compared to the Institutional setting, which has a correlation coefficient around 0.15.
Power

• .80 – chance of detecting a difference is 4 out of 5 if there is a real difference

• Most psychotherapy studies do not have sufficient power to detect small (r = .1) and medium (r = .3) effect sizes with power = .80
Participants Per Group Necessary to Detect Effect Sizes for an ANOVA w/ 2 Groups, Power = .80
Increasing Power

• Increase $N$
• Extreme conditions (e.g., severe psychopathology) may increase effects
  – Potent interventions may increase effects
• Use a pretest to account for within-subject variability and reduce error term
  – Without repeated measures: $ES = (m_1 - m_2)/s$
  – With repeated measures: $ES = (m_1 - m_2)/s\sqrt{1-r^2}$
• Decrease variability (error)
  – Heterogeneity of participants, inconsistent procedures, unreliable measures
Multiple Comparisons

• Bonferroni adjustment - Divide alpha (p = .05) by the number of comparisons
  – For 10 comparisons, p = .005 (.05/10)

• Multiple outcomes
  – Conceptually related as part of one domain
  – Empirically related - correlated

• Multivariate analyses – several measures in a single data analysis
  – Alpha adjustments for post hoc tests
Objections to Statistical Significance Testing

• All or none decision making ($p < .05$)
  – Studies without statistically significant differences are unlikely to be published
• Null hypothesis is rarely true
• Significance is a measure of N
  – Many psychology studies have low power
  – The same results in 2 studies may be statistically significant or nonsignificant
• Statistical significance says nothing about the importance of the effects
Objections to Statistical Significance Testing

• $p$ values refer to the likelihood that the finding would be obtained by chance if a large number of tests were run
  – They do not reflect the likelihood that the findings are true
  – A lower $p$ value ($p < .0001$) is not a stronger effect than a higher $p$ value ($p < .05$)

• A failure to detect a statistically significant difference does not necessarily mean that there is no difference
Interpretation of Results

• Overinterpretation
  – “Highly significant effect” has no statistical meaning
    • Nonsignificant trends generally should not be interpreted
    • “My results are not significant” – tests vs. results
  – “Predicts” can be used only if design warrants
  – “Fishing” – post hoc vs. a priori interpretations
  – “Implications” – how far are they from the actual results?

• Replication
  – 2 or more studies are better than one
No Differences: Reasons

• No differences exist
• Low power
• Manipulation failure
• Error
  – Participant heterogeneity
  – Inconsistent procedures
  – Unreliable measures
• Competing influences
  – Maturation
  – Statistical regression
No Differences

- May be informative if replicated
- Demonstrate the limits to the generalizability of the findings
Replication

- Is a finding reliable?
- Direct replication
- Systematic replication – varying features
  - Are there other IVs that may influence the DV?
- Replication by experimenters other than the original one
- Replication in non-laboratory contexts
  - Efficacy vs. effectiveness research