Major steps in the evaluation process

♦ Evaluation Design
♦ Measurement of the processes/outcomes
♦ Report of the evaluation results
♦ Interpretation of the evaluation results
Culturally Competent Evaluation

P.A. Olmos-Gallo
Lydia M. Prado
Mental Health Corporation of Denver

Evaluating Mental Health Performance: Let the Data Speak. Denver, CO. April, 2002
Evaluation Design

♦ No one is born culturally competent. We must learn it

♦ Develop specific cultural knowledge
  – Make sure that you include stakeholders of different ethnic backgrounds during the evaluation design

♦ Assure cultural competence of the evaluators/evaluation team
  – Knowledge of the people/community is VERY important; education may not be as much
Measurement of the processes/outcomes

♦ Make sure your instruments are culturally competent
  – Translation and back translation ARE A MUST on any culturally competent instrument

♦ Even if back-translated, we may face problems because the meaning may be different
  – “I feel blue” is not the same than “Me siento azul”, though the translation will be OK
How is cross-cultural equivalence manifested?

♦ Test scores from a given measure should be equally accurate for different groups
  – this means that reliability coefficients should be the same for all groups

♦ The factor structure on a given instrument should be the same for all relevant groups
  – e.g., if depression has 2 subdomains, it should have 2 subdomains for all cultural subgroups
equivalence -- cont’d

♦ Each item on a particular instrument should mean the same thing to people from different cultural groups
  – i.e., a psychological test that lacks item equivalence is in essence two different tests; one for each cultural group
What if we are not using questionnaires?

- Belief that etiology, expression, course, and outcome were universal and independent of cultural factors.

- Now, it is assumed that culture can play a role in psychopathology by:
  - Determining standards of normality
  - Creating personality configurations that may look like pathological in one culture but not in another
Even when using behavioral scores, we could face cross-cultural biases

- “Normal behaviors” in one culture can be classified as “pathological” in other
  - Dependency in Japan is valued, whereas in America it has negative connotations
  - Looking at the eyes to an authority figure will be considered “rude” in some cultures

- Culturally different individuals are not adequately represented in the norm groups
  - Classifying individuals of different cultures as pathological, based on norms may lead to tragedy
Use of behavior rating scales where rater and ratee come from different cultures
  – If cross-cultural differences result in biased ratings, then scales are not diagnostically valid for those groups
May suggest the need to generate norms for different cultures
Development of models to study the factor structure of the measure across different cultures
Techniques that can be used to test Instrument Equivalence

♦ Internal Consistency reliability estimates (Cronbach’s Alpha)
  – Checks if the instrument is consistently measuring what it was intended to measure (e.g., intelligence, mental functioning)
  – Comparisons across different ethnic groups
Techniques that can be used to test Instrument Equivalence (cont)

♦ **Factorial Invariance**
  
  – The main question it addresses: Do the items making a particular measuring instrument operate equivalently across different populations (e.g., White and Hispanics)?
  
  – Is the instrument measuring the SAME TRAITS in the SAME WAY across different ethnic groups?
    
    • The measurement model is group-invariant
Techniques that can be used to test Instrument Equivalence (cont)

• **Item equivalence**
  - A method for providing information about items on a particular instrument as well as scores for persons responding to the items
  - Historically, IRT has been used primarily by testing companies on aptitude and achievement tests
    - to determine how difficult the items are
    - to determine how well the items discriminate between people of high versus low ability
What does the Item equivalence tell us about rating scales?

♦ It provides “difficulty” indices for each item that indicate how relatively easy or difficult it is for a given individual to receive a high rating

  - on a clinical scale to measure some type of dysfunction, for example, an “easy” item suggests that even for clients with low levels of dysfunction, they have a high probability of receiving a high rating
  - a “difficult” item would require much higher levels of dysfunction in order to receive a high rating
Qualitative instruments can also be culturally incompetent

- We need to know what questions can be asked in different cultures
- We need to know how questions should be asked in different cultures
- We need to know “the adequate protocol” to ask questions
  - How to address people; Conventions about who is to speak when
Report of the Evaluation results

- Be aware of your own cultural bias
- Evaluators may be considered authority figures of in some cultures
  - Don’t over-use that power
- Some cultures have a strong oral tradition
  - Results can be presented as stories within the cultural context
  - Useful to have some visual aids if possible
- However, numbers are useful, if they are included in context
Interpretation of the Evaluation

♦ Be aware of your own cultural background when you are doing your interpretation
  - Interpretations are biased (“speaking to my long-dead ancestors” in some cultures does not mean schizophrenia)

♦ Do not overgeneralize. Outcomes in one program do not represent the culture

♦ Use multiple perspectives to examine and explain results
  - “What works for whom” rather than “what works”
Interpretation of the Evaluation (cont)

- Check alternative explanations to provide a more appropriate interpretation
- Check with your stakeholders if you are “getting it”
  - Member checking