Validity, Fairness and Testing

Methods of Investigating Fairness: Are They Doing Justice to Equity?

Kadriye Ercikan
University of British Columbia

Educational Assessment, Accountability and Equity Conversations on Validity Around the World Conference
ETS/Teachers College
Columbia University
March 28, 2012
Fairness as Consistency of Score Meaning

- Girls, Boys
- African American, Latino, White/Caucasian, Asian
- English Language Learners
- Students With Disabilities
- Countries in international assessments
Fairness Research Methodology

- Differential item functioning (DIF)
- Occurs when examinees from groups with similar ability levels perform differently
- Signals potential bias
- Is used to minimize bias in tests
Limitations in DIF Methodology

• DIF methodology assumes that the rest of the test is not biased

• Cannot determine bias without follow up investigations such as item reviews or cognitive studies

• Relies on **broad group definitions**
Heterogeneity in populations

• Diversity within groups
  • Gender (Cohen & Bolt, 2005)
  • African American/European American (Webb, Cohen & Schwanenflugel, 2008)
  • Countries (Oliveri, 2012)
  • Linguistic minority groups (Ercikan et al., 2012)
Average Percent Correct DIF Detection by Within Group Heterogeneity (Oliveri, 2012)
PISA 2006 Science Gender DIF (# DIF items out of 103 items) (Ercikan et al., 2012)
Final Note

• Our claims do not apply meaningfully to heterogeneous populations

• Leading to developing biased tests and threatening consistency of score meaning

• Reconsider our grouping definitions

• New methodologies with latent class modeling/mixture models approaches to capture heterogeneity