

**HINTS Data Users Conference**  
**Jan. 20-21, 2005 St. Pete Beach, FL**

**Evaluating Measurement Equivalence  
between Hispanic and Non-Hispanic  
Responders to the English Form of the  
HINTS Information SEeking Experience  
(ISEE) Scale**

**Bryce B. Reeve, Ph.D.**

**Neeraj K. Arora, Ph.D.**

**Outcomes Research Branch,  
Applied Research Program**

**Bryce's e-mail: [reeveb@mail.nih.gov](mailto:reeveb@mail.nih.gov)**



# Overview of Presentation(s)

- ▶ **Methodological Studies on Differential Item Functioning (DIF)**
  - ▶ Do groups respond differently to items within the HINTS because of:
    - “true” between-group differences on the measured construct (not DIF); or,
    - groups interpret an item differently resulting in biasing scores between groups (DIF).
- ▶ **Differential Item Functioning (DIF)**
  - ▶ What is it?
  - ▶ What are the implications for instruments containing DIF items?
  - ▶ What are some of the common methods to test for DIF?
  - ▶ How should we handle or control for DIF?
- ▶ **Illustrations of exploring DIF in the HINTS data**
  - ▶ Exploring DIF between Hispanic and Non-Hispanic respondents to the
    - Information Seeking Experience (ISEE) Scale. (Reeve)
    - Psychological Distress Scale. (Chang)

## The Challenge for Developing Culturally Sensitive Instruments

- ▶ A lot of care is taken when a survey is developed, adapted, or translated to different populations or groups.
- ▶ We hope our instruments are tapping into the same construct so that we may make across group comparisons.
- ▶ “Measurement Equivalence”

# Information SEeking Experience (ISEE) Scale

Based on the results of your overall search for information on cancer, tell me how much you agree or disagree with the following statements.

1. You wanted more information but did not know where to find it.
2. It took a lot of effort to get the information you needed.
3. You did not have the time to get all the information you needed.
4. You felt frustrated during your search for the information.
5. You were concerned about the quality of the information.
6. The information you found was too hard to understand.
7. You were satisfied with the information you found.

Would you say you: “Strongly Agree”, “Somewhat Agree”, “Somewhat Disagree”, or “Strongly Disagree”?

## The Challenge for Developing Culturally Sensitive Instruments

- ▶ However, populations may give culturally different responses to questions.
- ▶ The result is that one group may have higher scores than another group, not because they have higher levels of a trait but because of differences in their cultural beliefs.
- ▶ This is known as Differential Item Functioning (DIF) or item bias.

## DIF Study on ISEE Scale

- ▶ **Do Hispanics (n = 193) and Non-Hispanic whites (n = 2288) differentially respond to items in the ISEE scale?**
- ▶ **Do the items have culturally different meanings between the Hispanic and Non-Hispanic groups?**

## Definition: Differential Item Functioning

- ▶ **One group responds differently to an item than another group despite controlling for differences on the measured construct.**
- ▶ **Two respondents, from different populations but have equal levels of the underlying trait, have different probability of responding to an item**

## Impact: Differential Item Functioning

- ▶ **DIF items are a serious threat to the validity of the scale to measure the trait levels of members from different populations or groups.**
- ▶ **Scales containing such items may have reduced validity for between-group comparisons, because their scores may be indicative of a variety of attributes other than those the scale is intended to measure.**

## Classic DIF example from the literature

- ▶ **Azocar, Arean, Miranda, & Munoz (2001) found on the Beck's Depression Inventory:**
  - ▶ **Regardless of the level of depression, Hispanics are more likely to endorse "I feel like crying" than non-Hispanics.**
  - ▶ **Latino culture has practices and symbolisms that portray crying as an acceptable behavior reflecting suffering.**

## Quantitative Methods to Assess DIF

- ▶ **Classical Methods**
  - ▶▶ **Correlation and reliability analyses**
  - ▶▶ **Mantel-Haenszel chi-square method – contingency-table approach**  
(Holland & Thayer, 1988)
  - ▶▶ **Logistic Regression**  
(Swaminathan & Rogers, 1990)

## Quantitative Methods to Assess DIF

- ▶ **Structural Equation Modeling (SEM)**

- ▶▶ **Multi-group Analysis**

- ▶▶ **Multiple-Indicator/Multiple Cause (MIMIC) Models**

**(Fleischman, Spector, & Altman, 2002)**

## Quantitative Methods to Assess DIF

- ▶ **Item Response Theory (IRT) Modeling** (Embretson & Reise, 2000)
  - ▶▶ **Likelihood Ratio Tests**  
(Thissen, Steinberg, & Wainer, 1993).
  - ▶▶ **Signed and Unsigned Area Tests**  
(Raju, 1988, 1990).

## Item Response Theory (IRT) Modeling

- ▶ IRT models the relationship between a person's level on a latent variable (e.g., information seeking experience) and their likelihood of responding to each question in a scale (e.g., the ISEE)
- ▶ Item Parameter Invariance Feature
  - ▶▶ Item properties are invariant to group membership.
    - Difficulty or severity of the item
    - Relevance of the item to the underlying construct.
- ▶ If DIF is detected, IRT can control for item bias when estimating scores.

# DIF Analysis of the ISEE Scale

Controlling for the mean differences between Hispanics and Non-Hispanics (.25 of a standardized score), found DIF for...

1. You wanted more information but did not know where to find it.
2. It took a lot of effort to get the information you needed.
3. You did not have the time to get all the information you needed.
4. You felt frustrated during your search for the information.
5. **You were concerned about the quality of the information.**
6. The information you found was too hard to understand.
7. You were satisfied with the information you found.

The quality of the information on cancer was more important for non-Hispanic whites in the assessment of their information seeking experiences than Hispanics.

# Conclusions

- ▶ **Any evaluation of the psychometric properties of a questionnaire developed to measure a construct across two or more groups of importance to a study should include an assessment of DIF.**
  - ▶▶ **Language translations of an instrument (Azocar et al, 2001; Orlando & Marshall, 2002)**
  - ▶▶ **Racial and cultural groups (Morales, Reise, & Hays, 2000; Teresi, 2001)**
  - ▶▶ **Sex and age groups (Fleishman et al, 2002)**
  - ▶▶ **Risk and treatment groups. (Panter and Reeve, 2002)**
  - ▶▶ **Administration modes.**

# Conclusions

- ▶ **Quantitative Methods should co-exist with both qualitative and cognitive methods to build and revise instruments.**
- ▶ **While quantitative methods may detect DIF, it takes review by experts or cognitive interviewing with respondents to determine why an item is exhibiting DIF.**
- ▶ **What do you do with the DIF item?**
  - ▶ **Rewrite the item.**
  - ▶ **Remove the item.**
  - ▶ **Control for the underlying differences using an IRT model for scoring respondents.**

# References

- ▶ Azocar, F., Arean, P., Miranda, J., & Munoz, R.F. (2001). Differential item functioning in a Spanish translation of the Beck Depression Inventory. *Journal of Clinical Psychology*, 57(3), 355-365.
- ▶ Embretson, S.E., & Reise, S.P. (2000). *Item Response Theory for Psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- ▶ Fleishman, J.A., Spector, W.D., & Altman, B.M. (2002). Impact of differential item functioning on age and gender differences in functional disability. *Journal of Gerontology: Social Sciences*, 57B(5), S275-S284.
- ▶ Holland, P.W. & Thayer, D.T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H.I. Braun (eds.), *Test Validity* (p. 129-145). Hillsdale, NJ: Lawrence Erlbaum Associates
- ▶ Holland, P.W., & Wainer, H. (1993). *Differential Item Functioning*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- ▶ Morales, L.S., Reise, S.P., & Hays, R.D. (2000). Evaluating the equivalence of health care ratings by whites and Hispanics. *Medical Care*, 38(5), 517-527.
- ▶ Orlando, M., & Marshall, G.N. (2002). Differential item functioning in a Spanish translation of the PTSD checklist: detection and evaluation of impact. *Psychological Assessment*, 14(1), 50-59.
- ▶ Panter, A.T., & Reeve, B.B. (2002). Assessing tobacco beliefs among youth using item response theory models. *Drug and Alcohol Dependence*, 68 (supp. 1), 821-839.
- ▶ Raju, N.S. (1988). The area between two item characteristic curves. *Psychometrika*, 53, 495-502.
- ▶ Raju, N.S. (1990). Determining the significance of estimated signed and unsigned areas between two item response functions. *Applied Psychological Measurement*, 14, 197-207.
- ▶ Swaminathan, H., & Rogers, J.J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, 27, 361-370.
- ▶ Teresi, J.A. (2001). Statistical methods for examination of differential item functioning (DIF) with applications to cross-cultural measurement of functional, physical and mental health. *Journal of Mental Health and Aging*, 7(1), 31-40.
- ▶ Thissen, D., Steinberg, L., & Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models. In P.W. Holland & H. Wainer (Eds.) *Differential Item Functioning* (p. 67-114). Hillsdale, NJ: Lawrence Erlbaum Associates.