

# Comparability of Results across Years

**Richard Hill**  
**Center for Assessment**  
**CCSSO LSA Conference**  
**Nashville, TN—June 18, 2007**

# Basic Issue

---

- **Scores go up (or down) from one year to the next**
- **What might have caused that change?**
  - One possibility is a change (up or down) in the effectiveness of the educational system
  - Another possibility is that something else changed
  - What are those “something else?”

# Requirements for Comparability

---

- **Equivalent test**
- **Equivalent population**
- **Equivalent conditions**
- **Equivalent scoring**

# Equivalent Test

---

- **Same framework**
- **Same test specifications**
- **Solid equating**

# If the Framework Has Changed

---

- **Are the new questions measuring the same inherent skills as the old ones**
  - If yes, continue trend line
  - If not, restart trend line

# Determining If Changes Are Consequential

	Unique to Old 1 2 3 ... N	Common across Both N+1 N+2 ... N+M	Unique to New N+M+1 N+M+2 ...
Unique to Old 1 2 3 N			X
Common across Both	X		
Unique to New	X	X	

# Determining If Changes Are Consequential

---

- **Compute correlations of each item with each other item for all five cells**
- **Convert each correlation to a Fisher z**
  - $Z = .5 * [ \ln(1+r) - \ln(1-r) ]$
- **Compute mean and standard deviation within the five cells**
- **Compare means across the five cells (dividing by standard deviations to get effect size)**

# Determining If Changes Are Consequential

	Unique to Old	Common	Unique to New
Unique to Old	A	B	X
Common across Both	X	C	D
Unique to New	X	X	E



# Solid Equating

---

- All changes in total results are based *entirely* on changes in the equating items
- Were equating items presented identically?
  - Same item—no alterations
  - Same equivalent position
- Were equating items representative?

# Solid Equating (cont'd)

---

- **Was equating sample representative?**
- **Were any equating items deleting during the post-administration analysis?**

# Equivalence of Population

---

- **Did enrollments change?**
  - All students
  - Significant subpopulations
- **Was there a change in the percentage of students tested?**
  - All students
  - Significant subpopulations
- **High school dropouts**

# Equivalence of Conditions

---

## ■ **Stakes**

- “Stakes changes everything”

## ■ **Accommodation policies**

- Inclusion rules
- Accommodation rules

## ■ **Time**

- Testing time
- Time of testing
  - School schedules
  - Year-round schools

## ■ **Course-taking patterns**

# Equivalence of Conditions (Cont'd)

---

- **If concern, look at results district by district**

# Equivalence of Scoring

---

- **Constructed-response questions**
- **Method of arriving at a reported score**

# Random Fluctuation

---

- **Sampling error (small for most states)**
- **Standard error of equating (likely to be considerably larger)**
- **If error is larger than change, look for trends over a longer period of time**

# Thoughts? Suggestions?

---

- [rhill@nciea.org](mailto:rhill@nciea.org)
- Publication on [www.nciea.org](http://www.nciea.org)