An Index to Evaluate the Efficacy of Linking Using the Stocking-Lord Method

April 29, 2017
Introduction

In a setting using item response theory (IRT),

- Administered a set of equating items for two scales
- Linking
  - Assumed that the freely estimated item response functions (IRFs) for a set of equating items can be *linearly transformed* to their previously estimated on-scale IRFs *by a single common transformation function (SCTF)*.
  - Place two separately estimated item parameters to a single common scale
Introduction

The Stocking-Lord transformation (Stocking & Lord, 1983)

- Most commonly used in practice
- Produce more stable results rather than the moment methods (e.g., Baker & Al-Karni, 1991; Hanson & Béguin, 2002)
- Previous studies focused on ...
  - Comparing the characteristic curve methods with other competing linking methods (e.g., Kim & Cohen, 1992; Kim & Lee, 2006; Lee & Ban, 2010)
  - Developing a new linking method (e.g., He, 2013; Ragland, 2010)

Little research has evaluated the faithfulness of the Stoking-Lord linking results to the underlying SCTF assumption.
To check the performance of the transformation, generally simulation-based indices were used, other than linking coefficients:

- e.g.) Mean squared errors (MSE) of the estimated IRFs on the linked form

No index that can be used in operational settings to evaluate the efficacy of the transformation results.
Present Study

Purpose

- Develop an index to be used in operational settings to evaluate the SCTF assumption, with application to the Stocking-Lord method
- Evaluate the index via simulation studies and empirical analyses compared to other possible criteria
  - Type I error study – To show a proof of concept for the index
  - Power study – To explore the power of the proposed index in evaluating the resulting transformation when some items have drift
Simulation Study

Data Generation

- Use the flexMIRT (Cai, 2015)
- Generate Item response data to simulate two administrations (i.e., Year 1 and Year 2) of a set of common items

Simulation conditions

- Type I error study
  - 27 simulation conditions
  - 3 sample size × 9 ability distribution pair
- Power study
  - 80 simulation conditions
  - 2 sample size × 4 ability distribution pair × 5 magnitude of item parameter drift × 2 percentage of drifting items
Simulation Study

Data Generation (cont.)

- Item parameters
  - Type I error study
    - 20 equating item parameter estimates from an operational statewide test
    - Stable equating items without any item parameter drift (IPD)
    - $\bar{a} = .758$, $\bar{b} = -.211$, and $\bar{c} = .178$
  - Power study
    - 20 stable equating items + 7 (25%) drift items = 27 items in total
    - 20 stable equating items + 20 (50%) drift items = 40 items in total
    - 5 levels of magnitude of IPD: the magnitude of IPD was randomly selected from each of uniform distributions
    - Considered only unidirectional $b$-drift $\Rightarrow$ negative (-) sign of $b$-drift
Simulation Study

Data Generation (cont.)

- Ability parameters
  - Ability ($\theta$) distribution for Year 1 and Year 2 \sim normal distribution with different mean ($\mu$) and standard deviation ($\sigma$)
  - Type I error study
    - 9 conditions of a ability distribution pair
  - Power study
    - 4 conditions of a ability distribution pair
Simulation Study

Simulation Conditions

- Type I error study

<table>
<thead>
<tr>
<th>Factors</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Size</td>
<td>1,000, 2,000, 5,000</td>
</tr>
<tr>
<td>Difference in ability distribution between Year 1 and Year 2</td>
<td></td>
</tr>
<tr>
<td>N(0, 1) &amp; N(0, 1)</td>
<td></td>
</tr>
<tr>
<td>N(0, 1) &amp; N(0.25, 1)</td>
<td></td>
</tr>
<tr>
<td>N(0, 1) &amp; N(0.5, 1)</td>
<td></td>
</tr>
<tr>
<td>N(0, 1) &amp; N(0, 0.8^2)</td>
<td></td>
</tr>
<tr>
<td>N(0, 1) &amp; N(0, 1.25^2)</td>
<td></td>
</tr>
<tr>
<td>N(0, 1) &amp; N(0.25, 0.8^2)</td>
<td></td>
</tr>
<tr>
<td>N(0, 1) &amp; N(0.25, 1.25^2)</td>
<td></td>
</tr>
<tr>
<td>N(0, 1) &amp; N(0.5, 1.25^2)</td>
<td></td>
</tr>
<tr>
<td>N(0, 1) &amp; N(0.5, 0.8^2)</td>
<td></td>
</tr>
</tbody>
</table>
# Simulation Study

## Simulation Conditions

- **Power Study**

<table>
<thead>
<tr>
<th>Factors</th>
<th>Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size</td>
<td>1,000, 5,000</td>
</tr>
<tr>
<td>Difference in ability distribution between Year 1 and Year 2</td>
<td>N(0, 1) &amp; N(0, 1), N(0, 1) &amp; N(0.25, 1), N(0, 1) &amp; N(0, 1.25²), N(0, 1) &amp; N(0.25, 1.25²)</td>
</tr>
<tr>
<td>Percentage of drifting items (Total length of equating items)</td>
<td>25% (27 items), 50% (40 items)</td>
</tr>
<tr>
<td>Magnitude of IPD</td>
<td>Δb = 0, 0.20 ≤ Δb &lt; 0.40, 0.40 ≤ Δb &lt; 0.80, 0.80 ≤ Δb &lt; 1.20, 1.20 ≤ Δb &lt; 1.60</td>
</tr>
</tbody>
</table>
Simulation Study

Analysis Procedure

- **Step 1.** Item calibration
- **Step 2.** Scale transformation (Linking)
- **Step 3.** Evaluation (ind. HBI)
Simulation Study

Step 1. Item Calibration

- Separately estimated item parameters for the two testing occasions via marginal maximum likelihood estimation (MMLE)
- Use the flexMIRT® software (Cai, 2015)
- Three-parameter logistic (3PL) IRT model

\[
P_i(\theta_j) = c_i + \frac{1 - c_i}{1 + \exp\left(-a_i(\theta_j - b_i)\right)}
\]
Simulation Study

Step 2. Scale Transformation

- Use the Stocking-Lord method to estimate linking coefficients of A and B with their standard errors (SEs)
- Use the R package ‘equateIRT’ (Battauz, 2016)
- Stocking-Lord Transformation (Stocking & Lord, 1983)
  - Find linking coefficients that minimize the summation of the squared difference between the TCCs of equating items over examinees

\[
SL_{\text{diff}}(\theta_i) = \left[ \sum_{j:y} p_{ij}(\theta_{ji}; \hat{a}_{jj}, \hat{b}_{jj}, \hat{c}_{jj}) - \sum_{j:y} p_{ij} \left( \theta_{ji}; \frac{\hat{a}_{jj}}{A}, A\hat{b}_{jj} + B, \hat{c}_{jj} \right) \right]^2
\]

\[
SL_{\text{crit}} = \sum_{i} SL_{\text{diff}}(\theta_i)
\]
Simulation Study

Step 2. Scale Transformation (cont.)

- By the property of scale invariance to linear transformation, a linear function can make two separate IRT parameter estimates converted to the same common scale.

- New (Year 2) to old (Year 1) transformation
  - Item parameter estimates on the Year 2 scale are transformed to the common Year 1 scale by using a linear function:

\[
a_{i,y2 \ on \ y1} = \frac{a_{i,y2}}{A_i}
\]

\[
b_{i,y2 \ on \ y1} = A_i b_{i,y2} + B_i
\]
Simulation Study

Step 3. Evaluation Criteria

- Linking coefficients of A and B
  - Summarize linking coefficients using the mean across 500 replications

- Standard Errors (SEs) of A and B
  - To assess the magnitude of the sampling variation that affect the estimated linking coefficients
  - Empirical SEs
    - Standard deviations (SDs) of A and B across 500 replications
  - Asymptotic SEs
    - Computed based on the variance-covariance matrix of the item parameter estimates
    - Use the option in the ‘equateIRT’ package (Battauz, 2015) in R
Simulation Study

Step 3. Evaluation (cont.)

- Mean and SD of individual linking coefficients of A and B
  - Compute *individual linking coefficients for each item using every single item as an anchor*
  - Compute the mean and SD of A and B across all the equating items
  - Help us look at the variation in A and B

- Normalized root mean squared difference (RMSD)
  - Based on the weighted sum of squared distance between the TCCs
  - Divide the weighted sum by the number of common items

\[
\text{Normalized RMSD} = \frac{1}{m} \sqrt{\frac{\sum_{j=1}^{n} [TCC_{Y1}(\theta_j) - TCC_{Y2}^*(\theta_j)]^2 g(\theta_j)}{\sum_{j=1}^{n} g(\theta_j)}}
\]
Simulation Study

Step 3. Evaluation (cont.)

- Haebara-Based Index (HBI)
  - Based on the sum of root mean squared differences between the item response functions on the Year 1 and linked Year 2 scales
    \[ HBI = \sqrt{\frac{\sum_{i=1}^{m} [p_{iY1}(\theta_j) - p_{iY2}^*(\theta_j)]^2}{\sum_{j=1}^{n} g(\theta_j)}} } \]
  - On a 0 to 1 scale with an interpretation similar to the standardization DIF statistic (Dorans & Kulick, 1986)
  - \( HBI_{crit} = \text{critical value of } HBI \text{ at the level of } .05 \text{ and } .01 \text{ after getting a normal distribution of } HBI \text{ values over 500 replications} \)
  - Power rate = \( p(HBI > HBI_{crit}|HBI_{HA}) \) at the level of .05 and .01
Simulation Study

Results  |  Linking Coefficient A

- Type I error study
  - Slope (A) is affected by variability (SD) of the ability distribution, but not affected by sample size
Simulation Study

- Power study

Sample size 1,000 & 25% IPD

Sample size 5,000 & 25% IPD

Sample size 1,000 & 50% IPD

Sample size 5,000 & 50% IPD
Simulation Study

Results

- Type I error study
  - Intercept (B) is affected by the mean ability distribution, but not affected by sample size

![Graph showing the relationship between linking coefficient B and different mean ability distributions and sample sizes.]
Simulation Study

### Power Study

<table>
<thead>
<tr>
<th>Sample size</th>
<th>IPD</th>
<th>Delta</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,000</td>
<td>25%</td>
<td>0</td>
<td>0.50</td>
</tr>
<tr>
<td>1,000</td>
<td>25%</td>
<td>0.20</td>
<td>0.40</td>
</tr>
<tr>
<td>1,000</td>
<td>25%</td>
<td>0.40</td>
<td>0.30</td>
</tr>
<tr>
<td>1,000</td>
<td>25%</td>
<td>0.80</td>
<td>0.20</td>
</tr>
<tr>
<td>1,000</td>
<td>25%</td>
<td>1.20</td>
<td>0.10</td>
</tr>
<tr>
<td>1,000</td>
<td>50%</td>
<td>0</td>
<td>0.60</td>
</tr>
<tr>
<td>1,000</td>
<td>50%</td>
<td>0.20</td>
<td>0.50</td>
</tr>
<tr>
<td>1,000</td>
<td>50%</td>
<td>0.40</td>
<td>0.40</td>
</tr>
<tr>
<td>1,000</td>
<td>50%</td>
<td>0.80</td>
<td>0.30</td>
</tr>
<tr>
<td>1,000</td>
<td>50%</td>
<td>1.20</td>
<td>0.20</td>
</tr>
<tr>
<td>5,000</td>
<td>25%</td>
<td>0</td>
<td>0.60</td>
</tr>
<tr>
<td>5,000</td>
<td>25%</td>
<td>0.20</td>
<td>0.50</td>
</tr>
<tr>
<td>5,000</td>
<td>25%</td>
<td>0.40</td>
<td>0.40</td>
</tr>
<tr>
<td>5,000</td>
<td>25%</td>
<td>0.80</td>
<td>0.30</td>
</tr>
<tr>
<td>5,000</td>
<td>25%</td>
<td>1.20</td>
<td>0.20</td>
</tr>
<tr>
<td>5,000</td>
<td>50%</td>
<td>0</td>
<td>0.70</td>
</tr>
<tr>
<td>5,000</td>
<td>50%</td>
<td>0.20</td>
<td>0.60</td>
</tr>
<tr>
<td>5,000</td>
<td>50%</td>
<td>0.40</td>
<td>0.50</td>
</tr>
<tr>
<td>5,000</td>
<td>50%</td>
<td>0.80</td>
<td>0.40</td>
</tr>
<tr>
<td>5,000</td>
<td>50%</td>
<td>1.20</td>
<td>0.30</td>
</tr>
</tbody>
</table>
Simulation Study

Results  | Standard Error (SE) of A

- Type I error study
  - SE of A is affected by sample size and ability distribution
  - Empirical SE of A showed the similar pattern as asymptotic SE of A

![Empirical SE of A](image1)

![Asymptotic SE of A](image2)
Simulation Study

- Power study: Empirical SE of A (left) & Asymptotic SE of A (right)

Sample size 1,000 & 25% IPD

Sample size 5,000 & 25% IPD

Sample size 1,000 & 50% IPD

Sample size 5,000 & 50% IPD
Simulation Study

Results

- **Standard Error (SE) of B**

  - **Type I error study**
    - SE of B is affected by sample size and ability distribution
    - Empirical SE of B showed the *similar* pattern as asymptotic SE of B

![Empirical SE of B](image1)

![Asymptotic SE of B](image2)
Simulation Study

- Power study: Empirical SE of B (left) & Asymptotic SE of B (right)
Simulation Study

Results

Mean of Individual A and B

- Type I error study
  - The mean of individual A and B for a every single item across all equating items produced the *same* pattern as the regular A and B

![Graph showing Mean of individual A and B](image1)

![Graph showing Mean of Individual B](image2)
Simulation Study

- Power Study: Mean of individual A (left) and B (right)

Sample size 1,000 & 25% IPD

Sample size 5,000 & 25% IPD

Sample size 1,000 & 50% IPD

Sample size 5,000 & 50% IPD
Simulation Study

Results

- **SD of Individual A and B**
  - **Type I error study**
    - The average of SD of individual A and B for a single item across all equating items produced the similar pattern as the SE of the regular A and B, but its magnitude is different.
Simulation Study

- Power Study: SD of individual A (left) and B (right)

Sample size 1,000 & 25% IPD

Sample size 5,000 & 25% IPD

Sample size 1,000 & 50% IPD

Sample size 5,000 & 50% IPD
Simulation Study

Results

<table>
<thead>
<tr>
<th>Normalized RMSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type I error study</td>
</tr>
<tr>
<td>Y-axis = Normalized RMSD</td>
</tr>
<tr>
<td>Affected by sample size and variability of the ability distribution pair.</td>
</tr>
</tbody>
</table>

![Chart showing normalized RMSD values for different distributions and sample sizes.](chart)

- Sample size: 1000
- Sample size: 2000
- Sample size: 5000
Simulation Study

- **Power Study (Y-axis = Normalized RMSD)**

  - Sample size 1,000 & 25% IPD
  - Sample size 5,000 & 25% IPD
  - Sample size 1,000 & 50% IPD
  - Sample size 5,000 & 50% IPD
Simulation Study

Results

HBI (Haebara-Based Index)

- Type I error study (Y-axis = HBI)
  - The values of HBI looks quite consistent, but slightly affected by sample size due to IRT parameter estimation error
Simulation Study

- Type I error study (Y-axis = SD of HBI)
Simulation Study

- **Critical value of HBI** at the level of .05 and .01 under the H₀

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Type I Error</th>
<th>N(0, 1)</th>
<th>N(0.25, 1)</th>
<th>N(0.5, 1)</th>
<th>N(0, 0.8²)</th>
<th>N(0, 1.25²)</th>
<th>N(0.25,0.8²)</th>
<th>N(0.25,1.25²)</th>
<th>N(0.5,0.8²)</th>
<th>N(0.5,1.25²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,000</td>
<td>.05</td>
<td>0.0396</td>
<td>0.0400</td>
<td>0.0414</td>
<td>0.0427</td>
<td>0.0386</td>
<td><strong>0.0434</strong></td>
<td>0.0390</td>
<td>0.0430</td>
<td>0.0398</td>
</tr>
<tr>
<td></td>
<td>.01</td>
<td>0.0423</td>
<td>0.0426</td>
<td>0.0442</td>
<td>0.0457</td>
<td>0.0412</td>
<td><strong>0.0464</strong></td>
<td>0.0415</td>
<td>0.0461</td>
<td>0.0425</td>
</tr>
<tr>
<td>2,000</td>
<td>.05</td>
<td>0.0284</td>
<td>0.0288</td>
<td>0.0298</td>
<td>0.0303</td>
<td>0.0276</td>
<td><strong>0.0329</strong></td>
<td>0.0277</td>
<td>0.0329</td>
<td>0.0286</td>
</tr>
<tr>
<td></td>
<td>.01</td>
<td>0.0302</td>
<td>0.0307</td>
<td>0.0317</td>
<td>0.0322</td>
<td>0.0294</td>
<td><strong>0.0351</strong></td>
<td>0.0295</td>
<td>0.0351</td>
<td>0.0305</td>
</tr>
<tr>
<td>5,000</td>
<td>.05</td>
<td>0.0184</td>
<td>0.0188</td>
<td>0.0196</td>
<td>0.0198</td>
<td>0.0175</td>
<td><strong>0.0220</strong></td>
<td>0.0180</td>
<td>0.0220</td>
<td>0.0184</td>
</tr>
<tr>
<td></td>
<td>.01</td>
<td>0.0196</td>
<td>0.0201</td>
<td>0.0209</td>
<td>0.0211</td>
<td>0.0186</td>
<td><strong>0.0235</strong></td>
<td>0.0191</td>
<td>0.0235</td>
<td>0.0195</td>
</tr>
</tbody>
</table>

- At least 1,000 examinees, \( HBI_{\text{crit},a=0.05} = 0.043 \) and \( HBI_{\text{crit},a=0.01} = 0.046 \), respectively
- If \( HBI > HBI_{\text{crit}} \), the equating-item set is NOT working properly for the transformation
- **Note.** \( HBI_{\text{crit}} \) slightly decreases, as the sample size or the length of the equating items increases (b/c the estimation error)
Simulation Study

Power Study (Y-axis = HBI)

Sample size 1,000 & 25% IPD

Sample size 1,000 & 50% IPD

Sample size 5,000 & 25% IPD

Sample size 5,000 & 50% IPD
Simulation Study

- Power study (Y-axis = SD of HBI)

Sample size 1,000 & 25% IPD

Sample size 5,000 & 25% IPD

Sample size 1,000 & 50% IPD

Sample size 5,000 & 50% IPD
Simulation Study

- **Power rate**: $P(HBI > HBI_{crit} \mid HBI_{HA})$ at the level of .05 and .01

<table>
<thead>
<tr>
<th>% of IPD items</th>
<th>Sample size</th>
<th>Magnitude of IPD</th>
<th>N(0,1)</th>
<th>N(0.25,1)</th>
<th>N(0, 1.25^2)</th>
<th>N(0.25, 1.25^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>.05</td>
<td>.01</td>
<td>.05</td>
<td>.01</td>
</tr>
<tr>
<td>25%</td>
<td>1,000</td>
<td>0.2≤Δb&lt;0.4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.4≤Δb&lt;0.8</td>
<td>0.996</td>
<td>0.996</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.8≤Δb&lt;1.2</td>
<td>0.994</td>
<td>0.994</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.2≤Δb&lt;1.6</td>
<td>0.992</td>
<td>0.992</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5,000</td>
<td>0.2≤Δb&lt;0.4</td>
<td>0.996</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.4≤Δb&lt;0.8</td>
<td>0.998</td>
<td>0.998</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.8≤Δb&lt;1.2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.2≤Δb&lt;1.6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>50%</td>
<td>1,000</td>
<td>0.2≤Δb&lt;0.4</td>
<td>0.996</td>
<td>0.988</td>
<td>1</td>
<td>0.996</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.4≤Δb&lt;0.8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.8≤Δb&lt;1.2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.2≤Δb&lt;1.6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5,000</td>
<td>0.2≤Δb&lt;0.4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.4≤Δb&lt;0.8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.8≤Δb&lt;1.2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.2≤Δb&lt;1.6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Real Data Analysis

Purpose

- Demonstrate the performance of HBI in the real data sets
- Compare the different sets of equating items using the propose HBI
  - A set of all equating items
  - A final set of equating items from the operational selection
  - A final set of equating items selected by the iterative procedures (Jing, Louis, & Yu, 2017)
Real Data Analysis

Data

- An operational statewide test
- Science test in the 8th grade: 45 MC items with 8,663 examinees
- History test in high school: 60 MC items with 24,204 examinees

### Magnitude of Item Parameter Drift

<table>
<thead>
<tr>
<th></th>
<th>$\Delta a$</th>
<th>$\Delta b$</th>
<th>$\Delta c$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Science</td>
<td>.14</td>
<td>.12</td>
<td>.87</td>
</tr>
<tr>
<td>History</td>
<td>.13</td>
<td>.16</td>
<td>.31</td>
</tr>
</tbody>
</table>
Real Data Analysis

Analysis Procedure

- Analysis procedure is the same as the simulation
- Step 1. Item calibration — PARSACLE
- Step 2. Scale transformation — STUIRT
- Step 3. Evaluation — Linking coefficients, normalized RMSD, & HBI
# Real Data Analysis

## Results

<table>
<thead>
<tr>
<th></th>
<th>All equating items (45 items)</th>
<th>25 common items from a iterative procedure</th>
<th>17 common items from a operational selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1.1764</td>
<td>1.2807</td>
<td>0.9982</td>
</tr>
<tr>
<td>B</td>
<td>0.7654</td>
<td>0.9179</td>
<td>0.0003</td>
</tr>
<tr>
<td>Normalized RMSD</td>
<td>0.0037</td>
<td>0.0039</td>
<td>0.0048</td>
</tr>
<tr>
<td>HBI</td>
<td>0.1183</td>
<td>0.0829</td>
<td>0.0628</td>
</tr>
</tbody>
</table>

## Science Data

Comparison of TCCs using the linking items

![Graphs showing TCCs for Year 1 vs. Transformed Year 2](image1)

![Graphs showing TCCs for Year 1 vs. Transformed Year 2](image2)

![Graphs showing TCCs for Year 1 vs. Transformed Year 2](image3)
Real Data Analysis

Results  History Data

<table>
<thead>
<tr>
<th></th>
<th>All linking items (60 items)</th>
<th>22 common items from a iterative procedure</th>
<th>18 common items from a operational choice</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.9294</td>
<td>0.8893</td>
<td>0.9935</td>
</tr>
<tr>
<td>B</td>
<td>0.2943</td>
<td>0.2634</td>
<td>0.2173</td>
</tr>
<tr>
<td>RMSD</td>
<td>0.0039</td>
<td>0.0046</td>
<td>0.0040</td>
</tr>
<tr>
<td>HBI</td>
<td>0.0906</td>
<td>0.0254</td>
<td>0.0414</td>
</tr>
</tbody>
</table>

Comparison of TCCs using the linking items
The current study proposes an index – **Haebara-based index (HBI)** – to evaluate “whether the resulting transformation adequately adheres to the underlying assumption that the *same single linear transformation* satisfactorily transforms each item” to be applicable to *operational settings*

**Overall, the results demonstrate that ...**

- HBI is a good first step in this direction
- HBI shows significant promise in filling this gap in the monitoring of goodness-of-fit in operational administrations
Discussion

For Future Study ...

- Extend HBI to incorporate polychotomous items
- To obtain the estimated HBI standard error, taking into account IRT estimation error,
  - Use the “delta method”
  - Use the “simulation-based method” using directly the standard errors of item parameter estimates
- Investigate the impact of both the total number of linking items and the percent of drifting items in the linking set
Thank you.

juunhee@msu.edu