The teacher effect: Conceptualizing sensitivity as non-invariance

Megan Welsh
Jessica Flake
Yujia Li
Hariharan Swaminathan
Jerome D’Agostino

Instructional Sensitivity Conference
Lawrence, KS
11/14/13
Traditional approaches to identify sensitive items

- $O_1$: Pretest
- $X$: Instruction
- $O_2$: Posttest

Opportunity to Learn

Overall achievement

Demographics

Item performance
The solution: Multiple-group Measurement Invariance

Differences in item difficulties and/or discriminations across classrooms are assumed to be due to instruction.

Similar logic to value-added models where the difference between students’ observed and predicted scores are called a teacher effect.

Where \( v \) represents item difficulty and \( \lambda \) represents item discrimination.
The Challenge

- Measurement invariance studies typically require large sample sizes within each cluster.
- The number of parameters that we need to estimate is very large, especially relative to the number of participants.
A (preliminary) answer

Bayesian Structural Equation Modeling (BSEM)

- Newly available routines in Mplus 7.1 (Alignment Method) allow for Bayesian multiple-group measurement invariance studies for dichotomous variables.

- In simulation studies, it has reliably identified non-invariance in clusters of 100.
Our Work

• Examines how well this works in classroom-size clusters using data at two grade levels (3rd and 5th) from one school district.
  - 32 3rd grade classes with at least 20 students
  - 38 5th grade classes with at least 20 students

• Analyses focus on the state mathematics test
  - 44 multiple choice items on the 3rd grade test
  - 50 multiple choice items on the 5th grade test

• 23 district teachers (11 third-grade, 12 fifth-grade) also conducted a judgmental review of item sensitivity
Results

Model fit

• Undetermined: The number of parameters we are trying to estimate is greater than the number of observations.

• Could be investigated by a simulation study.

• We compared the item parameters generated using the full sample with those generated only for classes of at least 25 students and have found remarkably stable results.
**Results: Consistency Across Analyses**

Non-Invariant Parameters, Estimated using a Full and Partial Sample: 3rd Grade (616 parameters)

<table>
<thead>
<tr>
<th></th>
<th>Invariant</th>
<th>Non-Invariant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Sample</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Invariant</td>
<td>35.4%</td>
<td>1.0%</td>
</tr>
<tr>
<td>Non-Invariant</td>
<td>1.6%</td>
<td>62.0%</td>
</tr>
</tbody>
</table>

Non-Invariant Parameters, Estimated using a Full and Partial Sample: 5th Grade (800 parameters)

<table>
<thead>
<tr>
<th></th>
<th>Invariant</th>
<th>Non-Invariant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Sample</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Invariant</td>
<td>41.0%</td>
<td>1.4%</td>
</tr>
<tr>
<td>Non-Invariant</td>
<td>0.8%</td>
<td>56.9%</td>
</tr>
</tbody>
</table>
Results

Non-invariance

• A relatively small proportion of items were identified as non-invariant in any classroom.

• Items tended to be flagged for non-invariance in item difficulties.
3rd Grade Results: Frequency of Non-Invariant Items

- **44 items**
- **32 classrooms**
- **785 students**

No items exhibited non-invariance in both difficulty and discrimination.
5th Grade Results:
Frequency of Non-Invariant Items

- 50 items
- 38 classrooms
- 891 students

Proportion of Items with Classrooms Exhibiting Non-invariance in Difficulty

<table>
<thead>
<tr>
<th>Classrooms</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>50.0%</td>
</tr>
<tr>
<td>1</td>
<td>30.0%</td>
</tr>
<tr>
<td>2</td>
<td>12.0%</td>
</tr>
<tr>
<td>3</td>
<td>2.0%</td>
</tr>
<tr>
<td>4</td>
<td>2.0%</td>
</tr>
<tr>
<td>5</td>
<td>4.0%</td>
</tr>
</tbody>
</table>

Proportion of Items with Classrooms Exhibiting Non-invariance in Discrimination

<table>
<thead>
<tr>
<th>Classrooms</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>74.0%</td>
</tr>
<tr>
<td>1</td>
<td>22.0%</td>
</tr>
<tr>
<td>2</td>
<td>4.0%</td>
</tr>
</tbody>
</table>

Proportion of Items with Classrooms Exhibiting Non-invariance in Both Difficulty and Discrimination

<table>
<thead>
<tr>
<th>Classrooms</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>98.0%</td>
</tr>
<tr>
<td>1</td>
<td>2.0%</td>
</tr>
</tbody>
</table>
Next Steps...

- Examine items that were flagged
- Compare results with judgmental review
- Examine variance in item difficulty and discrimination across classrooms
- Replicate the BSEM analyses at the school level by randomly sampling schools in a statewide dataset
- Conduct a simulation study
Thanks!

Do you have any questions?

Megan Welsh
Department of Educational Psychology
Neag School of Education
megan.welsh@uconn.edu