Examining NAEP Mathematics with an Exploratory, Multilevel Item Factor Model

Nathan Dadey

University of Colorado Boulder

Gregory Camilli

Rutgers, The State University of New Jersey

Context

- Educational assessments are often used to compare teachers, schools, or **states**
- These comparisons are often limited
  - Subscales, meant to give information on content specific strengths and weaknesses, often do not provide information above and beyond the total score
  - E.g., on the 2009 National Assessment of Educational Progress (NAEP) in Fourth Grade Mathematics, the average state-level correlation, between the composite score and subscores is 0.95 (student-level is 0.85)

Context

- Alternative: Find subsets of items that states *do* differ on and use these items to define state-level dimensions
  - Camilli et al. (2008) took this approach by specifying, essentially, a multidimensional item response theory (MIRT) at the state-level

Literature

- Aggregate-Level IRT Modeling suggests value in exploring group-level dimensionality

(Camilli et al., 2008; Park & Bolt, 2008; Park & Lee, 2011)

- Work in Multilevel Factor Analysis suggests that examining total covariance matrices results in factor structures that are generally consistent with the within group structure, but not the between group structure

(Cronbach, 1979; Muthén, 1991; Van de Vijver & Poortinga, 2002; Muthén, 1994; Schweig; 2013)

- Here we extend this these two strands of work by using Full Information Item Factor Analysis

(Cai, 2010a, 2010b; Bock, Gibbons, & Muraki, 1988; Rijmen, 2009)

Research Questions

1. What impact does ignoring student-level multidimensionality have on state-level dimensionality?
2. What state-level dimensional structure is defensible? What are the interpretation(s) of the dimensions?
3. What relationships do the state-level dimensions have to the NAEP questionnaire variables?

NAEP 2009 4th Grade Assessment

- Nationally representative assessment, Multistage sampling design
- Dichotomized item response data for 159 Items
  - classified by (a) mathematical content area, (b) cognitive complexity and (c) format.
- Random sample of students within schools,
  - 84,608 students (within or level 1 unit) in
  - 7,900 schools in
  - 52 states/jurisdictions[1] (between or level 2 unit).

  drawn from the total NAEP sample of 197,700 students.

[1] All 50 states, Department of Defense schools and the District of Columbia.

Multilevel, Multidimensional Item Response Theory Model

- The probability of student *j* in state *k* answering dichotomous item *i* correctly, $X_{ijk} = 1$, is

$$P\left(X_{ijk} = 1 \mid \eta_{ijk}, b_i\right) = \frac{\exp[\eta_{ijk} + d_i]}{1 + \exp[\eta_{ijk} + d_i]}$$

where

d_i is the item intercept, and

η_{ijk} is a multilevel composite term.

[1]As implemented in the flexMIRT software v2.0 (Houts & Cai, 2012). Estimated using the Hastings Robbins-Monro Algorithm (Cai, 2010) using the overall student weight, ORIGWT.

The Composite Term

$$\eta_{ijk} = \mathbf{A}_i^C \boldsymbol{\theta}_k^C + \mathbf{A}_i^P \boldsymbol{\theta}_{jk}^P$$

State-level Slopes

State-level Scores (i.e., capacities)

Student-Level Slopes

Student-Level Scores (i.e., proficiencies)

Capacities, $\boldsymbol{\theta}_k^C$

- Meant to show states' strengths and weaknesses
- Dimensions that that may reflect the effects of
  - intentional policy decisions, or
  - population characteristics

Rotation

- Rotate the estimated state-level slopes, $\widehat{\mathbf{A}}^C$, and estimated student-level slopes, $\widehat{\mathbf{A}}^P$ (separately, using varimax)

RQ1: Ignoring Student-Level Multidimensionality

- Correlate the item slopes from
  - "1/5" solution
    - 1 student dimension, i.e., $\mathbf{A}_i^P$ is a 1×1 scalar
    - 5 state dimensions, i.e., $\mathbf{A}_i^C$ is a 1×5 vector of item slopes on the 5 state dimensions.
  - "5/5" solution.
- Conduct a targeted rotation on the 1/5 state-level slope matrix with the state-level 5/5 slope matrix as the target

RQ1: State-Level Item Slope Correlations

1/5 Solution (SSL)	5/5 Solution (SSL)				
	C1 (54.40)	C2 (3.31)	**C3 (14.42)**	C4 (1.42)	C5 (1.16)
C1 (28.48)	0.41	-0.38	0.78	0.26	-0.19
C2 (58.08)	0.98	-0.02	0.65	0.06	0.00
C3 (1.22)	0.02	-0.08	-0.35	0.96	0.24
C4 (5.71)	0.13	-0.95	0.02	0.12	0.00
C5 (2.22)	0.21	0.04	0.07	0.28	0.97

RQ1: State-Level Item Slope Correlations

1/5 Solution (SSL)	5/5 Solution (SSL)				
	C1 (54.40)	C2 (3.31)	**C3 (14.42)**	C4 (1.42)	C5 (1.16)
C1 (28.48)	0.41	-0.38	0.78	0.26	-0.19
C2 (58.08)	0.98	-0.02	0.65	0.06	0.00
C3 (1.22)	0.02	-0.08	-0.35	0.96	0.24
C4 (5.71)	0.13	-0.95	0.02	0.12	0.00
C5 (2.22)	0.21	0.04	0.07	0.28	0.97

RQ1: After Targeted Rotation

Rotated 1/5 Solution	5/5 Solution				
	C1	C2	C3	C4	C5
C1	0.99	0.14	0.63	-0.04	0.03
C2	0.14	0.99	0.06	-0.04	0.13
C3	0.63	0.06	0.99	-0.20	-0.14
C4	-0.05	-0.03	-0.21	1.00	0.26
C5	0.03	0.12	-0.14	0.26	1.00

RQ1: Conclusion

- Modeling student-level dimensionality changes pattern of loadings on the secondary state-level dimensions (under varimax rotation)

R2: Selecting a State-level structure

- Compare
  - 5/5,
  - 4/4,
  - 3/3, and
  - 2/2 solutions
- in terms of
  - cross-level measurement invariance, and
  - sums of squared loadings

R2: Cross-level Measurement Invariance

Assumes that

$$\mathbf{A}_i^C = \mathbf{A}_i^P$$

So

$$\eta_{ijk} = \mathbf{A}_i^C \boldsymbol{\theta}_k^C + \mathbf{A}_i^P \boldsymbol{\theta}_{jk}^P$$

$$= \mathbf{A}_{\boldsymbol{i}}(\boldsymbol{\theta}_k^C + \boldsymbol{\theta}_{jk}^P)$$

To check

$$\widehat{\mathbf{A}}_i^P \xrightarrow[\text{Targeted Rotation}]{} \widehat{\mathbf{A}}_i^C$$

R2: Results

- Cross-level measurement invariance does not hold, to a degree
  - Average diagonal correlation from the four student vs. state slope correlation matrices, after targeted rotation was, 0.65.
- Sum of Squared Loadings suggests 2 student-level factors and 2 state-level factors.

R2: 2/2 Solution

- At the state-level there was
  - a general dimension and
  - a dimension related to the ability to complete spatial patterns and make determinations related to probability.

Item	Loading	Content	Description
116	0.56	DAS	Probability of letter P
22	0.55	DAS	Determine probability
40	0.53	DAS	Find simple probability
78	0.49	GEO	Rabbit can't be fooled mirror image
32	0.48	GEO	Describe the position of a triangle
47	0.48	GEO	Recognize completed shape

Limitations & Considerations

- Alternative “tests” of Invariance and Selection Criteria
- Sampling Design (plausible values)
- Sample Size (52 states/jurisdictions)
- Alternative and Additional Approaches & Models
  - # of State Dimensions ≠ # of Student Dimensions
  - Profile Analysis
- Interpretation (conditional dimensional interpretations)