Background

• **My task:** Reflect on possibilities of improving self-reported attitude and background data, with a particular focus on ICCS

• **My orientation:** A quantitative methodologist with heavy leanings toward latent variable models

• **My perspective:** Acknowledging and allowing for heterogeneity in our models can help improve person-parameter estimation and cross-country comparisons on civic and citizenship scales
What is of interest in ICCS?

• Among other things, reporting, comparing and modeling:
  – Civic knowledge
  – Attitudes and perceptions
ICCS and other ILSAs

• We are often interested in this relationship

\[(\text{cognitive}) \text{ outcomes} = f[\text{background information}]\]

• This makes the \textbf{right side} of the equation as well as non-cognitive outcomes \textit{as important} as the cognitive outcomes – often where most time and efforts are spent
Producing scale scores

- Both types of scales rely on latent variable models and associated assumptions!
  - Dimensionality (usually, uni)
  - Latent variance explains all association in observed variables
  - Equivalence
- Plus, because of test design, civic knowledge requires some pretty heavy machinery to estimate proficiency
Estimation in a nutshell:

• Methods used to estimate population-level & sub-population level achievement:
  – Based on missing data methods;
  – Treats achievement as *unobserved* for every examinee;
  – Integrates IRT with a population model;
  – Uses what essentially amounts to a regression model to predict plausible achievement values for all examinees.
What does this *kind* of look like?

Background Data

IRT Model

θ

x₁ x₂ ... xₙ

δ₁ δ₂ δₙ

Conditioning Model

Too few item responses
The part I’m often interested in

\[ \theta = \Gamma y + e \]

- An important assumption: \( y \) is measured \textit{completely and without error}
Trouble in the conditioning model

• But is it reasonable to assume that, for example, a 4th grader will reliably know the answer to things like:

• Especially in economically developing countries, students answer very differently than their parents (Rutkowski & Rutkowski, 2010)
  – 34 non-OECD countries in 2012 PISA
Or that *anybody* consistently gets this…

In the box is a series of problems. Each requires you to understand a problem written in text and perform the appropriate calculations. Usually the problem talks about practical situations, but the numbers and people and places mentioned are made up. All the information you need is given. Here are two examples:

1) <Ann> is two years older than <Betty> and <Betty> is four times as old as <Sam>. When <Betty> is 30, how old is <Sam>?
2) Mr <Smith> bought a television and a bed. The television cost $625 but he got a 10% discount. The bed cost $200. He paid $20 for delivery. How much money did Mr <Smith> spend?

We want to know about your experience with these types of word problems at school.
Do not solve them!
(Please tick only one box in each row.)

<table>
<thead>
<tr>
<th></th>
<th>Frequently</th>
<th>Sometimes</th>
<th>Rarely</th>
<th>Never</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) How often have you encountered these types of problems in your <strong>mathematics lessons</strong>?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>b) How often have you encountered these types of problems in the tests you have taken at school?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>
And we know missing data are a problem

• PIRLS 2006:
  – Birthplace of student and both parents:
    • Germany (4.6%)
    • Indonesia (5.0%)
    • Israel (6.2%)
    • Kuwait (16.7%)
    • Qatar (5.1%)
    • South Africa (9.5%)
Consequences of error, in general

• From regression, we know that error in the independent variables causes problems:
  – Regression coefficients tend to be too small
  – Standard errors tend to be too big

• And these findings are from the simplest case
  – Tracking problems is harder in more complex models

• The conditioning model is a regression model, so we can expect the same problems
Consequences of missing data, in general

- Bias in parameter estimates
- Lost data
- Incorrect inferences
Results from my own work

• When background variables are noisy:
Results from my own work

- When background variables are missing (at random)
Country comparisons are important

• But this makes a strong assumption that a construct is understood and operationalized in the same way

• If this assumption is not met, it is never clear if differences on a scale are genuine latent mean difference or due to the instrument
To make comparisons, our measures should be equivalent

\[ \xi_1^{(1)} \]
\[ \xi_1^{(2)} \]

\[ \lambda_{21}^{(1)} \]
\[ \lambda_{31}^{(1)} \]
\[ \lambda_{21}^{(2)} \]
\[ \lambda_{31}^{(2)} \]

\[ x_1 \]
\[ x_2 \]
\[ x_3 \]
\[ x_1 \]
\[ x_2 \]
\[ x_3 \]

\[ \delta_1 \]
\[ \delta_2 \]
\[ \delta_3 \]
\[ \delta_1 \]
\[ \delta_2 \]
\[ \delta_3 \]

Plus the mean structure:

\[ \tau_i^{(1)} = \tau_i^{(2)} \quad \text{for all} \ i \]
Simple example

- Consider two populations with different relationships between an indicator $x$ and the latent variable, $\xi$

\[
x = 1.67 + 0.52LV
\]
\[
x = 2.35 + 0.70LV
\]
Simple example

• For two equal values of $\xi$ (say, 1), we get two different values of $x$ for country 1 and country 2:
  – Country 1: $2.35 + 0.70 \times 1 = 3.05$
  – Country 2: $1.67 + 0.52 \times 1 = 2.19$

• Since we only observe $x$, we can’t say whether differences in $x$ are due to real differences in $\xi$ or because of measurement differences

• And this is only one indicator. The problem is compounded when many variables comprise a scale
Equivalence evidence

• We know, from fairly extensive analyses of (at least) TALIS, that equivalence is not forthcoming.

• And given system-level heterogeneity, can we expect to achieve strong equivalence?
Allowing heterogeneity in our models

• Rather than ignoring and suppressing heterogeneity in our models, why not embrace it?

• Oliveri & von Davier (2011) recommend an approach that uses some country-specific item parameters

• In the last round of PISA, blocks of items for some countries were provided as an option – for countries whose proficiency was expected to be low
Oliveri & von Davier (2011)
PISA “Easy” Booklets
Rutkowski & Rutkowski (2013)

- Assuming scalar invariance of PISA items
  - Using OECD item parameters
    - Poor recovery of underlying proficiency for all countries
    - Outcome is worse for poor performing non-OECD countries
  - Allowing all countries to contribute to item parameter estimates
    - Poor recovery of underlying proficiency for all countries
    - Outcome is worse for high performing non-OECD countries
  - Allowing “easy” countries to contribute to parameter estimates
    - For easy booklet countries, even the OECD influence preserved recovery of underlying proficiency for those countries – surprisingly so.
Some combination is likely best

Group specific items

Unique parameters

December 10, 2013
What constraints

- Majority of items should be common, with *common* parameters
- A *few* items should be
  - Uniquely estimated
  - Unique to the group (geographic, cultural, etc)
- Group-unique items should be chosen to best capture cultural/linguistic/other nuances
At what cost & gain

• Scales should be longer, but probably not so much longer
  – This needs to be investigated

• Better recovery of underlying latent variables according to our limited investigations
  – That is what we are really, fundamentally after

• Doesn’t require much item development, with the exception of deciding on “groups” and relevant items

• And ICCS is already doing regional modules, so this is well underway in many aspects
Proficiency models to contextual models

- Much of this has been investigated with proficiency-type models (math, science, reading)
- So, the natural next step would be to apply this idea to the civic proficiency scales
- But, the models are no different at their core, so this should also hold for scales associated with
  - Antecedents
  - Processes
  - Outcomes
Finally

• Of course, this is just one possibility for strengthening scientific evidence
• There are very likely other viable means
• However, this approach would build on what substantive experts already know
• And it would also recognize and take account of the fact that there are cultural differences
ICCS Contextual Framework Model

Figure 3.1: Contexts for the development of learning outcomes related to civics and citizenship

Antecedents
- Wider community
- Educational system
- History and culture

School/classroom
- Characteristics
- Composition
- Resources

Student
- Characteristics

Home environment
- Family background
- Social group

Processes
- Wider community
- Educational policies
- Political events

School/classroom
- Instruction
- Governance

Student
- Socialization and learning

Home environment
- Communication
- Activities

Outcomes

Indicators related to:
- Civic society and systems
- Civic principles
- Civic participation
- Civic identities