Comment on NAEEd Report

A Focus on The Three Cs (Especially Common Sense)

Leslie Rutkowski
Associate Professor of Inquiry Methodology

April 12, 2018
First of all, thank you

- NCES
- NAEd
- Steering Committee
- Norwegian Research Council
- My research group in Oslo and Indiana
Overarching comments

• Very well-written survey of ILSAs
• Really liked
  – State of media and need for messaging/graphics
  – BICSE/National report idea
• This will be a reference text that I refer back to, especially with students, but also colleagues.
Cross-cultural studies as method

• Ref: Ellen Nolte
• Couldn’t agree more!
• This is where I’ll focus my comments
Massive (and ongoing) growth

• The report highlights significant growth in ILSAs over >2 decades
• PISA is growing most aggressively
  – So, I concentrate here
### A word on growth (and differences)

<table>
<thead>
<tr>
<th>Year</th>
<th>Number of OECD Countries</th>
<th>Avg GDP (in 2015 USD)</th>
<th>Number of Partner Countries</th>
<th>Avg GDP (in 2015 USD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015</td>
<td>34</td>
<td>$36,810</td>
<td>37</td>
<td>$15,149</td>
</tr>
<tr>
<td>2012</td>
<td>34</td>
<td>$41,819</td>
<td>31</td>
<td>$22,952</td>
</tr>
<tr>
<td>2009</td>
<td>34</td>
<td>$40,767</td>
<td>40</td>
<td>$17,856</td>
</tr>
<tr>
<td>2006</td>
<td>28</td>
<td>$39,836</td>
<td>27</td>
<td>$16,763</td>
</tr>
<tr>
<td>2003</td>
<td>30</td>
<td>$33,354</td>
<td>11</td>
<td>$18,212</td>
</tr>
<tr>
<td>2000</td>
<td>28</td>
<td>$27,965</td>
<td>15</td>
<td>$13,556</td>
</tr>
</tbody>
</table>
Achievement diversity (2015 science)
Test coverage and achievement levels

*Item distribution within PISA proficiency levels*

<table>
<thead>
<tr>
<th>Level</th>
<th>Score points on the PISA scale</th>
<th>2015 CBA</th>
<th>2015 PBA</th>
<th>Previous Cycles (PBA)</th>
<th>Total number of items (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>Higher than 707.93</td>
<td>9</td>
<td>3</td>
<td>1</td>
<td>13 (4.45%)</td>
</tr>
<tr>
<td>5</td>
<td>633.33</td>
<td>20</td>
<td>8</td>
<td>1</td>
<td>29 (9.93%)</td>
</tr>
<tr>
<td>4</td>
<td>558.73</td>
<td>51</td>
<td>20</td>
<td>4</td>
<td>75 (25.68%)</td>
</tr>
<tr>
<td>3</td>
<td>484.14</td>
<td>56</td>
<td>29</td>
<td>9</td>
<td>94 (32.19%)</td>
</tr>
<tr>
<td>2</td>
<td>409.54</td>
<td>36</td>
<td>21</td>
<td>6</td>
<td>63 (21.58%)</td>
</tr>
<tr>
<td>1a</td>
<td>334.94</td>
<td>10</td>
<td>3</td>
<td>2</td>
<td>15 (5.14%)</td>
</tr>
<tr>
<td>1b</td>
<td>260.54</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>3 (1.03%)</td>
</tr>
<tr>
<td>Below 1b</td>
<td>Less than 260.54</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0 (0%)</td>
</tr>
</tbody>
</table>
## Mapping to achievement levels

<table>
<thead>
<tr>
<th>Country</th>
<th>Mean</th>
<th>SD</th>
<th>Level 2 Upper Threshold</th>
<th>$p \leq L2$</th>
<th>Level 1a Upper Threshold</th>
<th>$p \leq L1a$</th>
<th>Level 1b Upper Threshold</th>
<th>$p \leq L1b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dominican Republic</td>
<td>332</td>
<td>72</td>
<td>484</td>
<td>.98</td>
<td>410</td>
<td>.86</td>
<td>335</td>
<td>.52</td>
</tr>
<tr>
<td>Algeria</td>
<td>376</td>
<td>69</td>
<td>484</td>
<td>.94</td>
<td>410</td>
<td>.69</td>
<td>335</td>
<td>.28</td>
</tr>
<tr>
<td>Tunisia</td>
<td>386</td>
<td>65</td>
<td>484</td>
<td>.93</td>
<td>410</td>
<td>.64</td>
<td>335</td>
<td>.21</td>
</tr>
<tr>
<td>Peru</td>
<td>397</td>
<td>77</td>
<td>484</td>
<td>.87</td>
<td>410</td>
<td>.57</td>
<td>335</td>
<td>.21</td>
</tr>
<tr>
<td>US</td>
<td>496</td>
<td>99</td>
<td>484</td>
<td>.45</td>
<td>410</td>
<td>.19</td>
<td>335</td>
<td>.05</td>
</tr>
<tr>
<td>Finland</td>
<td>531</td>
<td>96</td>
<td>484</td>
<td>.31</td>
<td>410</td>
<td>.10</td>
<td>335</td>
<td>.02</td>
</tr>
</tbody>
</table>
Wright maps

- Our Wright maps show the distribution of examinee proficiency on the $\theta$ continuum against the RP62 item location on the same continuum.

- Such a representation gives a clear picture of the degree to which a group of examinees are matched to a test.

- To further illustrate the points at hand, we also use the functional form of the IRT model to describe probabilities of correct answers for examinees along the continuum.
Devil’s advocate

• Dominican Republic is an outlier!

• Ok, but look who’s coming to dinner:
  • Cambodia
  • Guatemala
  • Panama
  • Senegal
  • Ecuador
  • Honduras
  • Paraguay
  • Paraguay
  • Zambia

PISA for Development

National Academy of Education

12 April 2018
A simulation

• We took Dominican Republic as our “population” – simulated data
• Gave Dominican Republic_{\text{sim}} PISA items
  • Repeated with successively easier tests
• Correlated estimated achievement with ESCS
# Simulation results

Overall achievement, conditional variance, and correlations

<table>
<thead>
<tr>
<th>Cond</th>
<th>Difficulty Mean</th>
<th>&quot;True&quot; Avg. Achievement</th>
<th>Est. Achievement</th>
<th>Conditional Variance</th>
<th>$\rho_{\theta,ESCS}$</th>
<th>$\hat{\rho}_{\theta,ESCS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.076</td>
<td>-1.648</td>
<td>-1.672</td>
<td>0.503</td>
<td>0.367</td>
<td>0.425</td>
</tr>
<tr>
<td>2</td>
<td>-0.735</td>
<td>-1.648</td>
<td>-1.656</td>
<td>0.472</td>
<td>0.367</td>
<td>0.400</td>
</tr>
<tr>
<td>3</td>
<td>-0.850</td>
<td>-1.648</td>
<td>-1.655</td>
<td>0.468</td>
<td>0.367</td>
<td>0.398</td>
</tr>
<tr>
<td>4</td>
<td>-0.939</td>
<td>-1.648</td>
<td>-1.655</td>
<td>0.466</td>
<td>0.367</td>
<td>0.396</td>
</tr>
<tr>
<td>5</td>
<td>-1.108</td>
<td>-1.648</td>
<td>-1.654</td>
<td>0.460</td>
<td>0.367</td>
<td>0.394</td>
</tr>
</tbody>
</table>
The promise of MST

• We can cover a lot more of the continuum with better precision:
A few challenges

• Item development for MST vs. trend
• The countries that would most benefit from very tailored items are countries where technology is a challenge
• A large number of easy items are necessary…
Another simulation

• We looked at how well “easy” booklets in PISA could recover proficiency in low performing countries
• We found that a lot of very easy items were needed to improve matters
Results

Easy cluster difficulty:

- △ -0.189
- + -0.689
- × -1.189
- * -1.689
- ○ c(-1.755, -1.545)
Moving forward

• Chapter 4: Promise of DBA
  • Yes! Along with MST – proposed in PSA – DBA is one tractable way to deal with big differences*
  • I disagree that DBA will have little effect on research and policy (see below)

• Chapter 5: ILSA for causal inference
  • A laudable goal
  • But until we get the measurement side sorted out…
Thank you!

Leslie Rutkowski
lrutkows@iu.edu
www.embracingheterogeneity.com