The Measurement Crisis

Ezequiel Molina *Economist, Education GP*
Julia Ladics *Consultant, Brown University*
In **Ecuador**, children scored \(~0.11\) standard deviations higher on math and language exams with high-quality teacher

*Similar findings in the USA, Uganda & Pakistan*

*Teachers* are the most important factor affecting learning in schools …

When students have a high quality teacher, they’re more likely to **attend college**, earn higher salaries, live in higher SES neighborhoods, and save more for retirement

*An average teacher can increase students lifetime income by up to **250K***
...but what determines a good teacher?

Observed characteristics give little indication

Good Teaching
To what extent the use of teacher practices scores are “valid”? 
The 5 C’s of Validation (Andrew Ho, 2018)

1. **Content** — Are these tools evidence based?

2. **Cognition** — How are these tools measuring these constructs? — aka RTFQ

3. **Coherence** — Are the scores derived from these tools reliable?

4. **Correlation** — Are the scores predictive of other variables?

5. **Consequence** — Do the use of the scores achieved the intended outcome?
Methodology – The Sample

Received 65 submissions from TTLs

42 tools are primary level government classroom observation tools

26 tools likely used to provide feedback to teachers

Able to gather reliability information on 16 tools

2 duplicate tools
11 tools for early childhood education, secondary, or college classroom observations
5 tools are school level or head teacher inspection tools with no classroom observation
5 submissions did not contain a classroom observation tool

4 tools used for non-governmental RCT only
8 tools used for program or project implementation
4 tools have never been used

Through communications with TTLs and local government officials, we were able to answer four key reliability questions 16 of the 26 tools
Methodology – The Framework

Classroom Culture
- Checks classroom environment
- Classroom management
- Positive social climate
- Equality and inclusiveness
- Family engagement
- Professionalism and respect

Instruction
- Feedback
- Checking for student understanding
- Content understanding
- Clear presentation and lesson structure
- Language development
- Lesson facilitation
- Critical thinking

Socioemotional Skills
- Student focus
- Motivating the classroom
- Discourse facilitation

43 Behaviors
15 Elements
3 Areas
1. Profile of the Average Tool
Profile of the **Average Tool**

**Ratings based**

**Used by Inspectors for 3 to 4 years**

**Measures 9 teaching practices**

<table>
<thead>
<tr>
<th>Rank</th>
<th>Element</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Lesson facilitation and discourse</td>
<td>16</td>
<td>100%</td>
</tr>
<tr>
<td>2</td>
<td>Checking for Student understanding</td>
<td>14</td>
<td>88%</td>
</tr>
<tr>
<td>3</td>
<td>Classroom management</td>
<td>13</td>
<td>81%</td>
</tr>
<tr>
<td>3</td>
<td>Content Understanding</td>
<td>13</td>
<td>81%</td>
</tr>
<tr>
<td>5</td>
<td>Checks classroom environment</td>
<td>12</td>
<td>75%</td>
</tr>
<tr>
<td>6</td>
<td>Clear Presentation and Lesson Structure</td>
<td>11</td>
<td>69%</td>
</tr>
<tr>
<td>6</td>
<td>Student focus</td>
<td>11</td>
<td>69%</td>
</tr>
<tr>
<td>8</td>
<td>Equality and Inclusiveness</td>
<td>10</td>
<td>63%</td>
</tr>
<tr>
<td>9</td>
<td>Positive Social climate</td>
<td>9</td>
<td>56%</td>
</tr>
<tr>
<td>9</td>
<td>Language Development</td>
<td>9</td>
<td>56%</td>
</tr>
</tbody>
</table>
2. Extent to Which Tools are Evidence-Based
Are Tools Evidence Based?
They are, in two ways...

The **Elements** they Measure

**How** they Measure
### Examples from Tools

**Checking for Student Understanding**

<table>
<thead>
<tr>
<th>Activity</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>5a. Students can follow simple instructions given by the teacher (e.g. to get their books out, form groups, put hands up)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5b. Students give appropriate responses to teacher questions (e.g. to comprehension questions when reading a story)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6e. Teacher asking comprehension questions when reading story/text</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Teacher asked questions to promote students’ critical thinking and check understanding.

1.5 Assessment of learning

<table>
<thead>
<tr>
<th>Scheme of work</th>
<th>Un satisfactory</th>
<th>Satisfactory</th>
<th>Good</th>
<th>Very good</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3. Extent to Which Tools are Reliable
Quantifying Reliability

Reliability Protocol Types

Are Tools Reliable?

1. Training
2. Observer Reliability
3. Evidence of Observer Reliability
4. Certification Exam
Quantifying Reliability

Reliability Protocol Types

Are Tools Reliable?

1. Training
2. Observer Reliability
3. Evidence of Observer Reliability
4. Certification Exam

16  5  1  2
Quantifying Reliability

Reliability Protocol Types

Are Tools Reliable?

1 of 4
10

2 of 4
4

3 of 4
2

4 of 4
0
The Measurement Crisis

Evidence-Based by Elements

16

Evidence-based by Measurement

2

- At least 3 of 4 reliability measures (1 tool)
- Less than 3 reliability measures (1 tool)

Not evidence-based by Measurement

14

- At least 3 of 4 reliability measures (1 tool)
- Less than 3 reliability measures (13 tools)
An Urgent Call to Action

More research on simpler tools attempting to capture quality of teacher practices (adding to the existing literature on time on task)

More effort on establishing validity evidence for the use of a score (There is NOT such as thing as validity of the tool)

Translate diagnostics into implementable programs
Thank you!

Ezequiel Molina, Economist, Education GP
Julia Ladics, Consultant, Brown University
Tracy Wilichowski, Analyst, Education GP
Noah Yarrow, Senior Education Specialist, Education GP

Please send comments and suggestions to molina@worldbank.org