

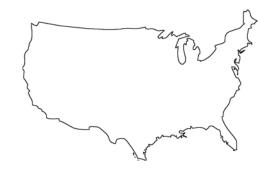
#### **The Measurement Crisis**

Ezequiel Molina *Economist, Education GP* Julia Ladics *Consultant, Brown University* 





#### Teachers are the most important factor affecting learning in schools ...



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



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

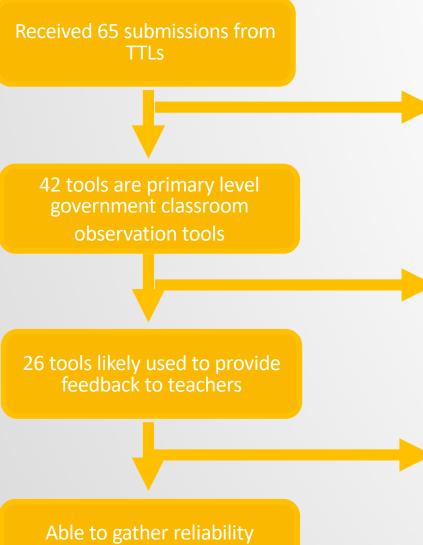
3. Coherence — Are the scores derived from these tools reliable?



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

5. Consequence — Do the use of the scores achieved the intended outcome?

### Methodology – The Sample



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

2 duplicate tools

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

information on 16 tools

## Methodology – The Framework



- Checks classroom environment
- Classroom management
- Equality and inclusiveness
- Professionalism and respect

- Checking for student understanding
- Content understanding
- Clear presentation and lesson structure
- Language development

- 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

| Rank | Element                                    | Frequency | Percentage |
|------|--|-----------|------------|
| 1    | Lesson facilitation and discourse          | 16        | 100%       |
| 2    | Checking for Student<br>understanding      | 14        | 88%        |
| 3    | Classroom management                       | 13        | 81%        |
| 3    | Content Understanding                      | 13        | 81%        |
| 5    | Checks classroom<br>environment            | 12        | 75%        |
| 6    | Clear Presentation and<br>Lesson Structure | 11        | 69%        |
| 6    | Student focus                              | 11        | 69%        |
| 8    | Equality and Inclusiveness                 | 10        | 63%        |
| 9    | Positive Social climate                    | 9         | 56%        |
| 9    | Language Development                       | 9         | 56%        |



#### 2. Extent to Which Tools are Evidence-Based



Are Tools Evidence Based? They are, in two ways...





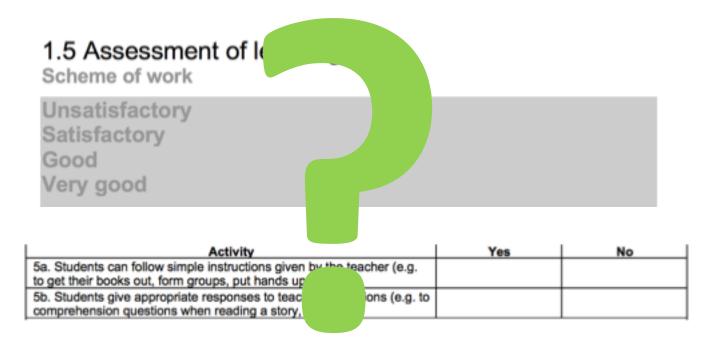
The Elements they Measure

How they Measure

## **Examples from Tools**

## Checking for Student Understanding

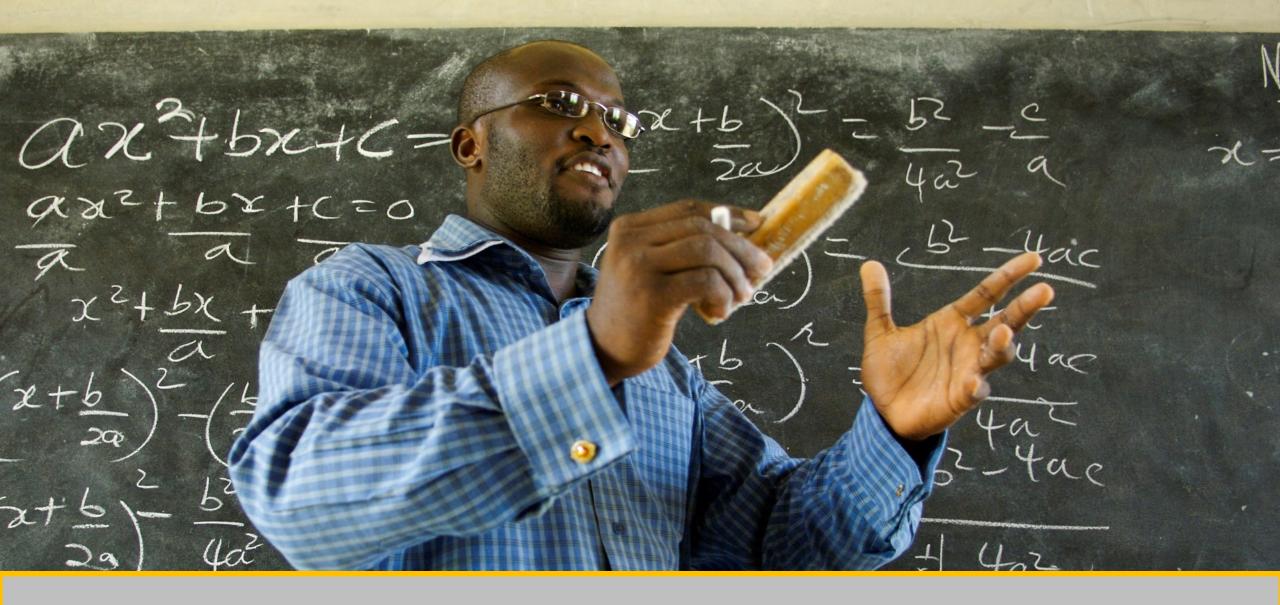
Teacher asked questions to promote students' critical thinking and check understanding



| 6e. Teacher asking comprehension questions when reading story/text |  |  |
|--|--|--|
|  |  |  |



use questioning appropriately to check students' understanding;



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

5

3. Evidence of Observer Reliability

#### 4. Certification Exam

16

Quantifying Reliability

**Reliability Protocol Types** 

#### Are Tools Reliable?



#### **The Measurement Crisis**



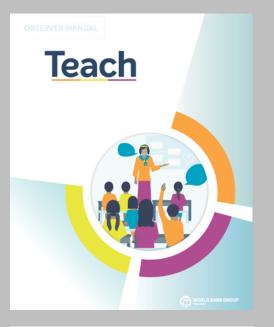


#### An Urgent Call to Action

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