The Predictive Validity of Clinical Practice Lessons: Experimental Evidence from Argentina

Alejandro J. Ganimian*, Andrew D. Ho* & Mariana Alfonso**
*Harvard University **Inter-American Development Bank

Society for Research on Educational Effectiveness
Washington, DC - Mar 6, 2015
A number of studies have shown that effective teachers matter.

Motivation

- The data that school systems collect about teachers at the time of hire does not predict their effectiveness.
- The data that researchers collect about explain only a small share of variation in teacher productivity.

- Graduate degrees (Hanushek, 1989; Hanushek et al., 2005; Hanushek & Rivkin, 2006; Rivkin et al., 2005)
- Subject-specific pedagogical knowledge (Gitomer, et al., 2014; Hill, et al., 2011)
- Socio-emotional skills (Duckworth, et al., 2009; Rockoff, et al., 2011)
There are 35+ alternative pathways into teaching across the world that require two weeks of clinical practice before entering teaching:
Prior studies suggest that, once teachers enter the school system, their effectiveness one year predicts their effectiveness on the next year.

Source: Kane et al. (2013)
The two questions that current research does not address are:

1) Can we **reliably distinguish between effective and ineffective teachers** during clinical practice lessons?

2) Can we use clinical practice lessons to **predict teaching effectiveness** during the school year?
All corps members of an alternative pathway into teaching in Argentina (Enseñá por Argentina or ExA) go through four stages:

- Stage 1: Online application
- Stage 2: Assessment center
- Stage 3: Summer training institute
- Stage 4: School year

- Individuals complete an online application
- Individuals participate in: group case study, interview, demonstration lesson, written exercise, critical thinking assessment
- Individuals participate in: teacher training workshops, clinical practice
- Teachers teach in schools
At each stage, ExA uses **different instruments**:

- **Stage 1:** Online application  
  - ExA scores applications using **structured rubrics**

- **Stage 2:** Assessment center  
  - ExA scores each activity in the assessment center using **structured rubrics**

- **Stage 3:** Summer training institute  
  - ExA scores clinical practice lessons using:  
    - student surveys  
    - classroom observations

- **Stage 4:** School year  
  - ExA scores school year lessons using:  
    - student surveys  
    - classroom observations  
    - principal surveys
At each stage, ExA measures different criteria:

- Stage 1: Online application
  - accomplishment
  - leadership
  - perseverance
Instruments

- At each stage, ExA measures **different criteria**:
  
  **Stage 2: Assessment center**
  
  **individual activities** (interview, demonstration lesson, written exercise, critical thinking assessment)
  
  **group activity** (group case study)
  
  - **leadership**
  - **communication**
  - **openness to feedback**
  - **perseverance**
  - **organization**
  - **communication**
  - **critical thinking**
At each stage, ExA measures **different criteria**:

- **Stage 3: Summer training institute**
- **classroom observations**
- **managing student behavior**
- **checking student understanding**
- **presenting content clearly**
- **implementing classroom procedures**
- **creating an environment for learning**
- **conveying importance of effort**
Stage 3: Summer training institute

At each stage, ExA measures different criteria:

- **student surveys**
- **conferring with students**
- **captivating students**
- **challenging students**
- **caring about students**
- **controlling discipline**
- **consolidating understanding**
- **clarifying concepts**

Instruments

- At each stage, ExA measures different criteria:
Instruments

At each stage, ExA measures different criteria:

- **Classroom observations**
- **Managing student behavior**
- **Checking student understanding**
- **Presenting content clearly**
- **Implementing classroom procedures**
- **Creating an environment for learning**
- **Conveying importance of effort**
- **Planning for every lesson**
- **Offering chances to practice**

Stage 4: School year
At each stage, ExA measures **different criteria**: managing student behavior, checking student understanding, presenting content clearly, tracking students' actions, implementing classroom procedures, creating an environment for learning, conveying importance of effort, tracking teachers' actions, planning for every lesson, offering chances to practice, assessing student progress, analyzing results.
At each stage, ExA measures **different criteria**:

- clarifying concepts
- controlling discipline
- consolidating understanding
- caring about students
- student surveys
- conferring with students
- captivating students
- challenging students
At all stages, teachers were randomly assigned to raters:

- Teacher #1 was blocked randomly assigned to rater #1.
- Teacher #1 was blocked randomly assigned to rater #2.
- Teacher #1 was blocked randomly assigned to rater #3.
- The raters were randomly assigned to match.
During clinical practice, teachers were also randomly assigned to students:
We observe **24 corps members** in the **2013** cohort at three stages and **32 corps members** in the **2014** cohort at all four stages:
1) Can the **online application** predict **teaching effectiveness**?

\[ Y_i = \phi_j + \gamma_k + \lambda_t + \beta X_i + \epsilon_{ijkt} \]

- **teaching effectiveness** (classroom observations + student surveys + principal surveys)
- **fixed effects** for blocks at stage 1
- **fixed effects** for blocks at stage 4
- **fixed effects** for cohorts
- **online application score** (structured rubrics)
The scores that teachers are assigned on their online application have a negative relationship with their teaching effectiveness.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stage 4</td>
<td>Stage 4</td>
<td>Stage 4</td>
<td>Stage 4</td>
<td>Stage 4 student surveys</td>
<td>Stage 4 classroom observations</td>
<td>Stage 4 principal surveys</td>
</tr>
<tr>
<td>Stage 1</td>
<td>-0.0001</td>
<td>-0.689**</td>
<td>-0.527**</td>
<td>-0.516**</td>
<td>-0.697*</td>
<td>-0.272</td>
<td>-0.476*</td>
</tr>
<tr>
<td>(0.181)</td>
<td>(0.257)</td>
<td>(0.222)</td>
<td>(0.221)</td>
<td>(0.358)</td>
<td>(0.278)</td>
<td>(0.266)</td>
<td></td>
</tr>
<tr>
<td>Stage 1 FEs?</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Stage 4 FEs?</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Cohort FEs?</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>N</td>
<td>47</td>
<td>47</td>
<td>47</td>
<td>47</td>
<td>52</td>
<td>54</td>
<td>51</td>
</tr>
</tbody>
</table>
This is not surprising, since these scores are used to select teachers:

Admitted applicants score better on average than non-admitted applicants.

Admitted applicants also vary less in their performance than non-admitted applicants.

It is possible that there is little variation left across admitted applicants.
2) Can the assessment center predict teaching effectiveness?

\[ Y_i = \phi_j + \gamma_k + \lambda_t + \beta X_i + \epsilon_{ijkt} \]

- **teaching effectiveness** (classroom observations + student surveys + principal surveys)
- **fixed effects** for blocks at stage 2
- **fixed effects** for cohorts
- **assessment center score** (structured rubrics)
- **fixed effects** for blocks at stage 4
Results

There is no clear relationship between the scores assigned to teachers on the *assessment center* and their teaching effectiveness.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 2 FEs?</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Stage 4 FEs?</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Cohort FEs?</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>N</td>
<td>47</td>
<td>47</td>
<td>47</td>
<td>47</td>
<td>52</td>
<td>54</td>
<td>51</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stage</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 4</td>
<td>-0.012</td>
<td>0.169</td>
<td>0.114</td>
<td>0.122</td>
<td>0.090</td>
<td>0.311</td>
<td>-0.083</td>
</tr>
<tr>
<td></td>
<td>(0.141)</td>
<td>(0.207)</td>
<td>(0.211)</td>
<td>(0.229)</td>
<td>(0.211)</td>
<td>(0.261)</td>
<td>(0.215)</td>
</tr>
</tbody>
</table>
If we only consider the scores on the **individual activities** in the assessment center, the coefficient is **consistently estimated around zero**.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 4</td>
<td>0.074</td>
<td>0.023</td>
<td>0.0001</td>
<td>0.021</td>
<td>-0.059</td>
<td>0.177</td>
<td>0.064</td>
</tr>
<tr>
<td></td>
<td>(0.104)</td>
<td>(0.118)</td>
<td>(0.108)</td>
<td>(0.113)</td>
<td>(0.165)</td>
<td>(0.128)</td>
<td>(0.156)</td>
</tr>
</tbody>
</table>

Stage 2 FEs?  ✔  ✔  ✔  ✔  ✔  ✔  ✔
Stage 4 FEs?  ✔  ✔  ✔  ✔  ✔  ✔  ✔
Cohort FEs?  ✔  ✔  ✔  ✔  ✔  ✔  ✔
N             47  47  47  47  52  54  51
Results

- This is not surprising, since these scores are used to select teachers:

Admitted applicants **perform better** in individual activities, but **do not vary less** than non-admitted applicants.

Admitted applicants **perform better and vary less** in the group activity than non-admitted applicants.
If we consider the scores on the **individual activities** in the assessment center, the relationship is **negative, but not statistically significant**.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stage 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Stage 4 student surveys</td>
<td>Stage 4 classroom observations</td>
<td>Stage 4 principal surveys</td>
</tr>
<tr>
<td>group activity</td>
<td>-0.096 (0.108)</td>
<td>-0.101 (0.120)</td>
<td>-0.107 (0.119)</td>
<td>-0.103 (0.122)</td>
<td>-0.265 (0.173)</td>
<td>-0.037 (0.142)</td>
<td>-0.055 (0.157)</td>
</tr>
<tr>
<td>Stage 2 FEs?</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Stage 4 FEs?</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Cohort FEs?</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>N</td>
<td>47</td>
<td>47</td>
<td>47</td>
<td>47</td>
<td>52</td>
<td>54</td>
<td>51</td>
</tr>
</tbody>
</table>
The demonstration lessons in the assessment center predict teachers’ scores in the classroom observations during the school year.

<table>
<thead>
<tr>
<th></th>
<th>()</th>
<th>()</th>
<th>()</th>
<th>()</th>
<th>()</th>
<th>()</th>
<th>()</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>demo lesson</td>
<td>0.158</td>
<td>0.122</td>
<td>0.116</td>
<td>0.127</td>
<td>0.101</td>
<td>0.300**</td>
<td>0.097</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage 2 FEs?</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Stage 4 FEs?</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Cohort FEs?</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>N</td>
<td>47</td>
<td>47</td>
<td>47</td>
<td>47</td>
<td>52</td>
<td>54</td>
<td>51</td>
</tr>
</tbody>
</table>

The demonstration lessons in the assessment center predict teachers’ scores in the classroom observations during the school year. Results
3) Can clinical practice lessons predict teaching effectiveness?

\[ Y_i = \phi_j + \gamma_k + \beta X_i + \epsilon_{ijk} \]

- **Teaching effectiveness** (classroom observations + student surveys + principal surveys)
- **Fixed effects** for blocks at stage 3
- **Clinical practice score** (classroom observations + student surveys)
- **Fixed effects** for blocks at stage 4
Results

Teachers’ score on clinical practice predicts their effectiveness in classroom observations and principal surveys during the school year.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 4</td>
<td>0.284</td>
<td>0.338</td>
<td>0.435</td>
<td>Stage 4 student surveys</td>
<td>Stage 4 classroom observations</td>
<td>Stage 4 principal surveys</td>
</tr>
<tr>
<td>Stage 3 FEs?</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Stage 4 FEs?</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>N</td>
<td>26</td>
<td>26</td>
<td>26</td>
<td>29</td>
<td>31</td>
<td>29</td>
</tr>
</tbody>
</table>

(0.216)  (0.370)  (0.376)  (0.631)  (0.364)  (0.427)
The relationship between **student surveys** during clinical practice and effectiveness is **positive, but not statistically significant**.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 4</td>
<td></td>
<td></td>
<td></td>
<td>Stage 4 student surveys</td>
<td>Stage 4 classroom observations</td>
<td>Stage 4 principal surveys</td>
</tr>
<tr>
<td>Stage 3 student surveys</td>
<td>0.215 (0.156)</td>
<td>0.149 (0.231)</td>
<td>0.043 (0.215)</td>
<td>0.238 (0.344)</td>
<td>0.280 (0.259)</td>
<td>0.117 (0.269)</td>
</tr>
</tbody>
</table>

Stage 3 FEs? ✔ ✔ ✔ ✔ ✔ ✔

Stage 4 FEs? ✔ ✔ ✔ ✔ ✔ ✔

N 26 26 26 29 31 29
The same is true for **classroom observations** during clinical practice.

<table>
<thead>
<tr>
<th>Stage 3 classroom observations</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.073</td>
<td>0.076</td>
<td>0.158</td>
<td>0.001</td>
<td>0.412*</td>
<td>0.399</td>
<td></td>
</tr>
<tr>
<td>(0.152)</td>
<td>(0.196)</td>
<td>(0.179)</td>
<td>(0.315)</td>
<td>(0.228)</td>
<td>(0.249)</td>
<td></td>
</tr>
</tbody>
</table>

Stage 3 FEs? ✔ ✔ ✔ ✔ ✔ ✔

Stage 4 FEs? ✔ ✔ ✔ ✔ ✔ ✔

N 26 26 26 29 31 29
4) Can clinical practice lessons predict teaching effectiveness, after accounting for teachers’ performance in application and selection?

\[ Y_i = \phi_j + \gamma_k + \beta_1 X_{i1}^1 + \beta_2 X_{i2}^2 + \beta_3 X_{i3}^3 + \epsilon_{ijk} \]

- fixed effects for blocks at stage 3
- online application score (structured rubrics)
- assessment center score (structured rubrics)
- clinical practice score (classroom observations + student surveys)
Results

- Clinical practice lessons predict effectiveness even when we hold constant scores on the online application and assessment center.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stage 4</td>
<td>Stage 4</td>
<td>Stage 4</td>
<td>Stage 4 student surveys</td>
<td>Stage 4 classroom observations</td>
<td>Stage 4 principal surveys</td>
<td></td>
</tr>
<tr>
<td>Stage 3</td>
<td>0.294</td>
<td>0.462</td>
<td>0.493</td>
<td>0.442</td>
<td>1.089**</td>
<td>0.872**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.243)</td>
<td>(0.405)</td>
<td>(0.384)</td>
<td>(0.546)</td>
<td>(0.380)</td>
<td>(0.349)</td>
<td></td>
</tr>
<tr>
<td>Stage 1?</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Stage 2?</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Stage 3 FEs?</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Stage 4 FEs?</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>26</td>
<td>26</td>
<td>26</td>
<td>29</td>
<td>31</td>
<td>29</td>
<td></td>
</tr>
</tbody>
</table>
Results

› We do not find **heterogeneity by cohort** (2013 y 2014).

› We also do not find **heterogeneity by timing of effectiveness information** (start v. late in the school year).

› We only find **heterogeneity by performance quintile** between stages 3 and 4.
The **predictive validity of clinical practice lessons** is stronger among teachers who performed in the bottom quintile.
Conclusions

- In this context, **clinical practice lessons** show promise to **predict differences in teaching effectiveness**.
  - This lessons **add information not captured** by the **application and selection** processes.
  - They are **most useful to distinguish among** teachers who performed poorly during clinical practice.
  - They are **most useful** to predict performance on the **classroom observations** and **principal surveys**.

- The **application and selection scores** do not predict differences between program admits.
  - However, **these variables were used to select teachers**, so it is possible that there is too little variation left to predict effectiveness.
Many thanks!

Email: alejandro_ganimian@mail.harvard.edu
Website: http://scholar.harvard.edu/alejandro_ganimian