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
I. Motivation

» A number of studies have shown that effective teachers matter

I. 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.

(Hanushek, 1989; Hanushek et al., 2005; Hanushek & Rivkin, 2006; Rivkin et al., 2005)

(Gitomer, et al., 2014; Hill, et al., 2011)

(Duckworth, et al., 2009; Rockoff, et al., 2011)

teacher certification

graduate degrees

subject-specific pedagogical knowledge

socio-emotional skills
I. Motivation

» 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)
I. Motivation

» There are 35+ alternative pathways into teaching across the world that require two weeks of clinical practice before entering teaching:
I. Motivation

» Growing number of teacher preparation programs require trainees to practice teaching

» Teacher residency programs (Silva, McKie, Knechtel, Gleason & Makowsky, 2014)

» Coaching-based teacher training programs (Gross & DeArmond, 2011; Kraft & Blazar, 2014)

If practice lessons predict teacher performance, we could use them to inform decisions about certification, selection, allocation, support and mentoring
II. Research Questions

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?
III. Context

» All corps members of an alternative pathway into teaching in Argentina (Enseñá por Argentina or ExA) go through four stages:

- **Stage 1: Online application**
  - Individuals complete an online application

- **Stage 2: Assessment center**
  - Individuals participate in:
    - group case study
    - interview
    - demonstration lesson
    - written exercise
    - critical thinking assessment

- **Stage 3: Summer training institute**
  - Individuals participate in:
    - teacher training workshops
    - clinical practice

- **Stage 4: School year**
  - Teachers teach in schools
III. Context

• 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**
IV. Instruments

» At each stage, ExA measures different criteria:
IV. 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**
- **Critical thinking**
IV. 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
IV. Instruments

» At each stage, ExA measures **different criteria**:

- **Stage 3: Summer training institute**
  - Student surveys
  - Conferring with students
  - Captivating students
  - Controlling discipline
  - Caring about students
  - Clarifying concepts
  - Consolidating understanding
  - Challenging students
IV. Instruments

» At each stage, ExA measures different criteria:
IV. Instruments

» At each stage, ExA measures different criteria:

Stage 4: School year

principal surveys

- 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
IV. Instruments

» At each stage, ExA measures different criteria:
V. Assignment

» At all stages, teachers were randomly assigned to raters:
V. Assignment

During clinical practice, teachers were also randomly assigned to students:
VI. Sample

» We observe 24 corps members in the 2013 cohort at three stages and 32 corps members in the 2014 cohort at all four stages:
VII. Empirical Strategy and Results

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)
VII. Empirical Strategy and Results

The scores that teachers get 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>Stage 4</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></td>
<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>
</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>
VII. Empirical Strategy and Results

- 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.

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

Graphs by pexa and se1_cohort
VII. Empirical Strategy and Results

2) Can the assessment center predict teaching effectiveness?

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

- fixed effects for blocks at stage 2
- fixed effects for cohorts
- assessment center score (structured rubrics)

**teaching effectiveness** (classroom observations + student surveys + principal surveys)
VII. Empirical Strategy and Results

» There is no clear relationship between the scores teachers receive 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</td>
<td>Stage 4</td>
<td>Stage 4</td>
<td>Stage 4</td>
<td>Stage 4</td>
<td>Stage 4</td>
<td>Stage 4</td>
<td>Stage 4</td>
</tr>
<tr>
<td></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>
<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>
VII. Empirical Strategy and Results

» 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 2</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>individual activities</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>
<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>
VII. Empirical Strategy and Results

» If we consider the scores on the group 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>Stage 2</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>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>
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.

This is not surprising, since these scores are used to select teachers.
## VII. Empirical Strategy and Results

» The demonstration lessons in the assessment center predict teachers’ scores in the classroom observations during the school year.

<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 demo lesson</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 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>
VII. Empirical Strategy and Results

3) Can the 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
- **fixed effects** for blocks at stage 4
- **clinical practice score** (classroom observations + student surveys)
VII. Empirical Strategy and Results

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

<table>
<thead>
<tr>
<th>Stage 3</th>
<th>Stage 4</th>
<th>Stage 4</th>
<th>Stage 4 student surveys</th>
<th>Stage 4 classroom observations</th>
<th>Stage 4 principal surveys</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(5)</td>
<td>(6)</td>
</tr>
<tr>
<td>Stage 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.284</td>
<td>0.338</td>
<td>0.435</td>
<td>0.401</td>
<td>1.060**</td>
</tr>
<tr>
<td></td>
<td>(0.216)</td>
<td>(0.370)</td>
<td>(0.376)</td>
<td>(0.631)</td>
<td>(0.364)</td>
</tr>
<tr>
<td>Stage 3 FEs?</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>
</tr>
<tr>
<td>N</td>
<td>26</td>
<td>26</td>
<td>26</td>
<td>29</td>
<td>31</td>
</tr>
</tbody>
</table>
VII. Empirical Strategy and Results

» 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 3 student surveys</td>
<td>0.215</td>
<td>0.149</td>
<td>0.043</td>
<td>0.238</td>
<td>0.280</td>
<td>0.117</td>
</tr>
<tr>
<td></td>
<td>(0.156)</td>
<td>(0.231)</td>
<td>(0.215)</td>
<td>(0.344)</td>
<td>(0.259)</td>
<td>(0.269)</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>
VII. Empirical Strategy and Results

» The same is true for classroom observations during clinical practice.

<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 3 classroom observations</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></td>
<td>0.073</td>
<td>0.076</td>
<td>0.158</td>
<td>0.001</td>
<td>0.412*</td>
<td>0.399</td>
</tr>
<tr>
<td></td>
<td>(0.152)</td>
<td>(0.196)</td>
<td>(0.179)</td>
<td>(0.315)</td>
<td>(0.228)</td>
<td>(0.249)</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>
VII. Empirical Strategy and Results

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

\[ Y_i = \phi_j + \gamma_k + \beta_1 X_i^1 + \beta_2 X_i^2 + \beta_3 X_i^3 + \epsilon_{ijk} \]

- **fixed effects** for blocks at stage 3
- **teaching effectiveness** (classroom observations + student surveys + principal surveys)
- **online application score** (structured rubrics)
- **clinical practice score** (classroom observations + student surveys)
- **assessment center score** (structured rubrics)
## VII. Empirical Strategy and 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>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 3</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 3</td>
<td>0.294 (0.243)</td>
<td>0.462 (0.405)</td>
<td>0.493 (0.384)</td>
<td>0.442 (0.546)</td>
<td>1.089** (0.380)</td>
<td>0.872** (0.349)</td>
</tr>
<tr>
<td>Stage 1?</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>
</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>
VII. Empirical Strategy and Results

» We do not find heterogeneity by cohort (2013 vs. 2014).

» We also do not find heterogeneity by the timing of the effectiveness measurement (early vs. late in the school year).

» We only find heterogeneity by performance quintile between stages 3 and 4.
VII. Empirical Strategy and Results

» The **predictive validity of clinical practice lessons** is **stronger** among teachers who performed in the **bottom quintile**.

![Graph showing the relationship between different stages of development and their predictive validity.](image-url)
VIII. Conclusions

» In this context, **clinical practice lessons** show promise to **predict differences in teaching effectiveness**
  » These 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.
Thank you!

marianaa@iadb.org

http://www.iadb.org/en/topics/education/