How effective are computer-based teacher training programs? Evidence from a randomized controlled trial in El Salvador

Aymo Brunetti¹, Konstantin Büchel¹, Martina Jakob², Ben Jann², Christoph Kühnhanß², Daniel Steffen¹

¹Department of Economics, University of Bern
²Institute of Sociology, University of Bern

Seminar on «Analytical Sociology: Theory and Empirical Applications», Venice International University, November 18–21, 2019
1. Motivation

2. A look back: Results from the CAL-IMPACT project in 2018 (RCT on computer-assisted learning in primary schools in El Salvador)

3. Pilot study on content knowledge of primary-school teachers in El Salvador

4. First results of an RCT on computer-assisted teacher training in 2019
Motivation
THE PROBLEM: LEARNING CRISIS

• Improved school enrollment rates in developing countries ...
• ... but poor learning outcomes
• „Schooling is not Learning“

➢ „Learning Crisis“ (World Bank)
THE PROBLEM: 2ND GRADE MATH QUESTIONS

Figure: Percent of correct answers on second grade math questions, Source: Baseline data collected in February 2018 (N=3,532)
Example: \( 45 \div 9 = \underline{\hspace{2cm}} \)

Correct answers:

- 3rd gr. \( 3\% \)
- 4th gr. \( 9\% \)
- 5th gr. \( 28\% \)
- 6th gr. \( 39\% \)
THE PROBLEM: BASELINE TEST SCORES

What's 8 : 2?  What time is it?
CAL-IMPACT: INTERVENTIONS (ADDITIONAL MATH LESSONS)

- 2 x 90 min./week
  40 classes, ≈ 800 children

- 2 x 90 min./week
  39 classes, ≈ 800 children

- 2 x 90 min./week
  30 classes, ≈ 800 children
CAL-IMPACT: Design (Implementation by www.consciente.ch)

Baseline Test

Control 1: Control classes in treatment schools (40 classes)
Control 2: Control classes from different schools (40 classes)

T 1: Additional CAL-classes with teacher (2x90min/week)
T 2: Additional CAL-classes with supervisor (2x90min/week)
T 3: Additional traditional classes (2x90min/week)

Endline Test

5.5 Months

Baseline Test

Control 1: Control classes in treatment schools (40 classes)
Control 2: Control classes from different schools (40 classes)

T 1: Additional CAL-classes with teacher (2x90min/week)
T 2: Additional CAL-classes with supervisor (2x90min/week)
T 3: Additional traditional classes (2x90min/week)

Endline Test

5.5 Months
CAL-IMPACT: RESULTS

\[ \beta_1 \beta_3 = 0, \ p-value: 0.13 \]

- **CAL + Teacher**: p-value: 0.07*
- **CAL + Supervisor**: p-value: 0.19
- **Teacher**: p-value: 0.63

Lower bound: control units within treatment schools
CAL-IMPACT: RESULTS

![Graph showing additional math knowledge (in school years) for CAL + Teacher, CAL + Supervisor, and Teacher with p-values.](image)

- **CAL + Teacher**: p-value: 0.07*
- **CAL + Supervisor**: p-value: 0.00*
- **Teacher**: p-value: 0.01*

*Lower bound: control units within treatment schools
Upper bound: control units from control schools*
CAL-IMPACT: LESSONS LEARNED

• CAL instructed by teachers has the largest impact.

• (Weak) evidence that CAL is more effective than additional lessons taught by teachers.

• Strong spillover effects.

• As a byproduct of the project, we noticed that teachers’ knowledge of the content they were supposed to teach was really poor, therefore ...
PART III
TEACHER TESTS 2018
Teacher Tests 2018: Design

- Random sample of 224 primary-school math teachers in El Salvador (Department of Morazan)
- Math test covering topics taught in 2nd to 6th grade
Teacher Tests 2018: Results

N = 224; survey design taken into account
Teacher Tests 2018: Results

Number sense & arithmetics
- Two digit addition
- Two digit subtraction
- Four digit by two digit division
- Compute percentage
- Rewrite addition as multiplication
- Add two fractions

Geometry & measurement
- Convert meters to kilometers
- Compute area of rectangle
- Convert seconds to hours
- Compute volume of cube
- Compute area of circle

Data & statistics
- Represent information in graph
- Interpret descriptive graph

Percent of correct answers

N = 224; survey design taken into account


**Table 2:** Relation between teacher’s test score and students’ learning over a eight month evaluation period

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Years-of-schooling effect</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>grade-specific score (in 10 PP)</td>
<td>0.131**</td>
<td>0.125**</td>
<td>0.133**</td>
<td>0.155***</td>
<td>0.146*</td>
</tr>
<tr>
<td></td>
<td>(0.042)</td>
<td>(0.043)</td>
<td>(0.039)</td>
<td>(0.037)</td>
<td>(0.057)</td>
</tr>
<tr>
<td>overall score (in 10 PP)</td>
<td>0.124**</td>
<td>0.117**</td>
<td>0.130***</td>
<td>0.151***</td>
<td>0.159**</td>
</tr>
<tr>
<td></td>
<td>(0.038)</td>
<td>(0.039)</td>
<td>(0.036)</td>
<td>(0.033)</td>
<td>(0.058)</td>
</tr>
<tr>
<td><strong>Standardized learning effect</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>grade-specific score (std.)</td>
<td>0.093**</td>
<td>0.088**</td>
<td>0.095**</td>
<td>0.111***</td>
<td>0.103*</td>
</tr>
<tr>
<td></td>
<td>(0.032)</td>
<td>(0.032)</td>
<td>(0.029)</td>
<td>(0.028)</td>
<td>(0.043)</td>
</tr>
<tr>
<td>overall score (std.)</td>
<td>0.098**</td>
<td>0.092**</td>
<td>0.102**</td>
<td>0.121***</td>
<td>0.125*</td>
</tr>
<tr>
<td></td>
<td>(0.031)</td>
<td>(0.032)</td>
<td>(0.030)</td>
<td>(0.028)</td>
<td>(0.048)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class level controls</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>School level controls</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Teacher controls</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>School fixed effects</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Standard errors in parentheses (clustered by schools).
All models include controls for grade and CAL treatment assignment.
Number of observations: 2765 students, 119 teachers, 48 schools.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
PART IV
COMPUTER-ASSISTED TEACHER TRAINING 2019
In-service teacher training program to ...

• improve teacher content knowledge in math
• to improve their teaching,
• and, hopefully, to improve student math skills

Treatment (incentivized):

• self-studying using computer-assisted learning software
• participation in four workshops (problems solving, recapitulation)

Implementation:

• Using Kolibri with Khan Academy contents
• In cooperation with NGO Consciente (www.consciente.ch)
CATT: Sample/Design

Population: primary school math teachers in Morazán

Sample: 313 teachers from 175 different schools applied for participation in the study
selection of the worst performing teacher of every school

Randomization: 87 teachers in the treatment group
88 teachers control group
stratified by baseline test scores and gender

Balance: almost identical baseline test results and balanced in variables such as gender, experience, and education

Attrition: no endline test for 11 teachers (6%)
Feb 2019: Registration phase

Mar 2019: Baseline teacher tests and surveys

April 2019: Start of intervention

Intervention with participants

Aug/Sep 2019: End of intervention & endline teacher tests

Jan 2020: Start of new school year

Trained participants teach students in new school year

Sep/Oct 2020: Endline student tests
CATT: Preliminary Results

Control group

Treatment group

Percent correct

Baseline test

Endline test
CATT: Preliminary Results

**Control group**

- Baseline test: 45%
- Endline test: 50%

**Treatment group**

- Baseline test: 45%
- Endline test: 55%
CATT: PRELIMINARY RESULTS

Frequency
-20 -15 -10 -5 0 5 10 15 20 25 30 35 40
endline test score – baseline test score

control group
treatment group
CATT: PRELIMINARY RESULTS

Effect in percentage points

Standardized effect

Treatment effect

- excluding controls
- including controls
CATT: Preliminary Results

The chart illustrates the treatment effect (in percentage points) across different domains and levels. The domains include arithmetics, geometry, and data and statistics. The levels are divided into grade 2/3, grade 4, grade 5, and grade 6. The horizontal bars represent the range of treatment effects observed within each domain and level.
CATT: Preliminary Results

Gender (p=0.06)
- Male
- Female

Age (p=0.02)
- 23-39
- 40-49
- 50-62

Treatment effect (percentage points)
Next year we will see whether teachers’ knowledge gain translates into better learning outcomes among students.