Expanding School Time and the Value of Computer-Assisted Learning*

Lessons from a Randomized Experiment in El Salvador

Konstantin Büchel[a,b], Martina Jakob[c,d],
Christoph Kühnhanss[c,d], Daniel Steffen[a,b], Aymo Brunetti[a,b]

[a]Center for Regional Economic Development, University of Bern
[b]Department of Economics, University of Bern
[c]Department of Sociology, University of Bern
[d]Consciente – Unterstützungsverein für El Salvador, Bern

Evaluation Design
(The evaluation design is registered on the *AEA RCT Registry* and the summary of the registration is attached to this file)

April 27, 2018

*We would like to thank Michael Gerfin (U. Bern), Ben Jann (U. Bern), Florian Keller (SKBF, Aarau), Ulf Liebe (U.Bern), Blaise Melly (U. Bern), Urs Moser (U. Zurich), Mauricio Romero (UC San Diego), Erik Snowberg (Caltech), and the participants at the Nadel Workshop (ETH Zürich) and Brown Bag Seminar (Department of Economics, University of Bern) for helpful comments and invaluable feedback.

1 Introduction

Education is a major driver of economic development and a determinant of social stability. Investment in human capital is thus largely seen as a sustainable strategy to improve conditions in the developing world. While enrolment rates have dramatically increased in the last few decades, education outcomes are still comparatively poor in many low-income countries (e.g. OECD, 2016). Therefore, the key question to be answered is, how the quality of education can be improved in developing countries.

By engaging students more actively in learning and tailoring content to their prior knowledge, technology has the potential to improve the quality of teaching. Computer-assisted learning (CAL) is less sensitive to the performance of overstrained teachers and allows for self-paced and interactive learning. Even in large and heterogeneous classes – a typical setting for low-income countries – students can thus work at their own pace and get instant feedback. Albeit promising, research on the impact of software-based teaching in developing countries is still in its early stages and geographically concentrated on Southeastern Asia. It is unclear how the existing evidence is applicable to Latin American countries, as the software used in earlier studies are not available in Spanish.[1] This contribution evaluates the impact of a CAL-project using the software *Khan Academy*, which is available in eleven languages including Spanish, Portuguese, French and English.[2]

The Salvadorian government has recently shown considerable effort in improving not only the accessibility but also the quality of education. The new policy named "Integrated System of Inclusive Full Time Schools" constitutes an attempt to expand school days to the afternoon, and to promote learning through innovative teaching and open and flexible curricula. In the context of this countrywide programme, the Salvadorian Ministry of education

[1] See Snilstveit et al. (2015) for a systematic review of schooling interventions in developing countries. While the review lists studies on technology-based teaching in Latin America, the cited studies do not focus on software assisted maths lessons but large scale provision of hardware in public schools.

[2] A list on the available languages supported by Khan Academy can be found on `https://khanacademy.zendesk.com/hc/en-us/articles/202483750-Is-Khan-Academy-available-in-other-languages-` (last accessed 14.01.2018).

cooperates with NGOs to collectively address the serious learning deficiencies of students in public schools (MINED, 2013). In this spirit, the NGO Consciente will implement a CAL-project in the rural district of Morazán, El Salvador, which aims at improving numeracy skills of primary school pupils. The intervention is conducted in coordination with the Salvadorian Ministry of Education and is intended to be a pilot project for a potential large scale programme to improve students' skills through technology-based teaching.

The evaluation of this project will allow us to (1) assess the effect of CAL on students' numeracy skills, (2) to appraise whether the effect is mainly attributable to additional maths classes or the use of computers, and (3) to compare the cost-effectiveness of different versions of the project. Besides providing valuable scientific evidence on the usefulness of software-based learning in developing countries, the results will serve as basis for the NGO's scaling strategy for this intervention.

2 Intervention and Timeline

The schooling intervention implemented by Consciente consists of three different treatments (see Figure 1). All treatments are additional maths classes in the afternoon using different schooling inputs. The additional classes are offered in the afternoon to complement regular classes that are held in the morning.

Each treatment consists of two additional lessons of 90 minutes per week. The first treatment offers additional maths classes conducted by contract teachers. They are hired by Consciente and teach refresher courses that repeat the maths curricula of lower grades. The second and the third treatment are additional maths classes based on computer-assisted learning (CAL) technology. Students work with the software Khan Academy, which allows them to learn independently and – more important – at their own level and pace. The difference between these two latter treatments is that the additional computer classes are either conducted by a temporarily contracted supervisor or by a temporarily contracted

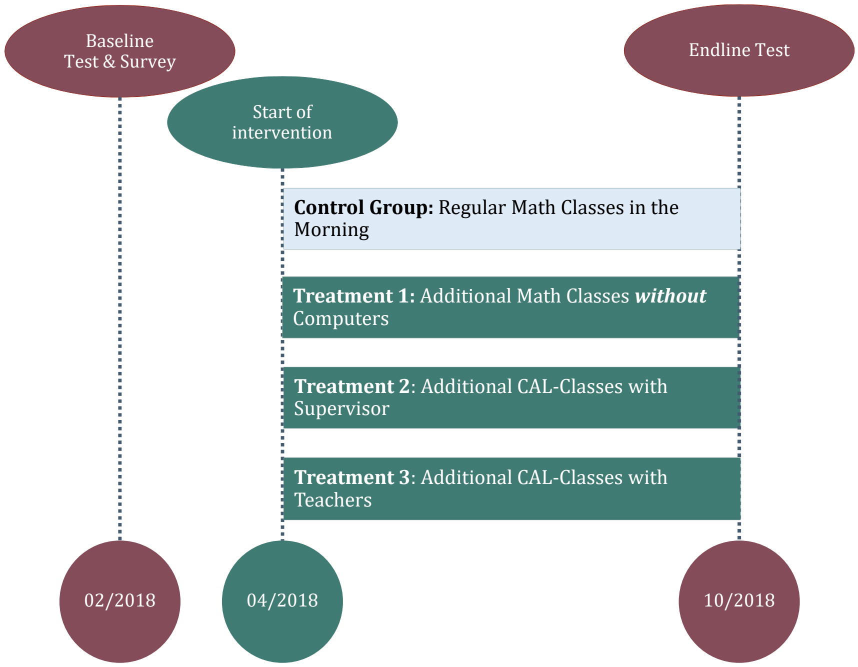

Figure 1: Timeline and the Three Intervention Arms

maths teacher. *Supervisors* provide technical support but should not assist with questions regarding maths. *Teachers*, in contrast, are also allowed to explain mathematical concepts to students, although students should mainly work independently with the computer programme. Beside these three treatment groups, control classes are also part of the evaluation: They do not get any treatment but serve as counterfactual to identify the causal impact of the three treatments on the students' numeracy skills (see Section 4).

3 Data on Main Outcome: Numeracy Skills

Our main outcome of interest are the students' numeracy skills. To measure numeracy skills, we conduct three standardized maths tests during the school year 2018 (see Figure 1). The standardized maths tests are designed as follows:

1. Summarizing the Salvadorian curriculum (grades 1–6) along the three topics (a.) number sense & arithmetic, (b.) geometry & measurement, and (c.) data & probability.

2. Mapping of test items on the Salvadorian curriculum. We use the following sources for test items: (a.) official text books of El Salvador, (b.) publicly available items from the STAR[3] evaluations in California, (c.) publicly available items from the VERA[4] evaluations in Germany, and (d.) exercises from the Swiss textbook MATHWELT.

3. Gathering of pilot data on 180 test items answered by 600 Salvadorian pupils in October 2017; estimating the difficulty and discrimination parameters of test questions based on *Item Response Theory* (e.g. de Ayala, 2008).

4. Design of (paper and pencil) maths tests using insights from step 3. The items are selected such that they reflect the weighting in the official curriculum: 60–65% number sense & arithmetic, 30% geometry & measurement, 5–10% data & probability.

4 Identification Strategy

A naive approach to estimate the intervention's impact would simply rely on changes in test scores across time without comparing them to a control group. Obviously, that would be problematic because maths skills of pupils improve during the school year even absent the NGO's work. Hence, the plain difference between baseline and endline test scores among beneficiaries (henceforward: treatment group) would overestimate the true causal effect of the intervention.

A more plausible approach compares the test score changes in the treatment group to test score changes in a control group, which serves the evaluator as counterfactual.[5] Assuming that – absent the intervention – treatment and control individuals are on average identical, any systematic differences in maths skills observed at the endline can then be attributed

[3]Further information on the Standardized Testing and Reporting (STAR) programme in California is available online: `www.cde.ca.gov/re/pr/star.asp` (last accessed: 14.01.2018).

[4]VERA is coordinated by the Institut für Qualitätsentwicklung im Bildungswesen (IQB), see `www.iqb.hu-berlin.de/vera` (last accessed: 14.01.2018).

[5]The true (but never observable) counterfactual is the (hypothetical) outcome of the treatment group, if they had not be part of the NGO's intervention.

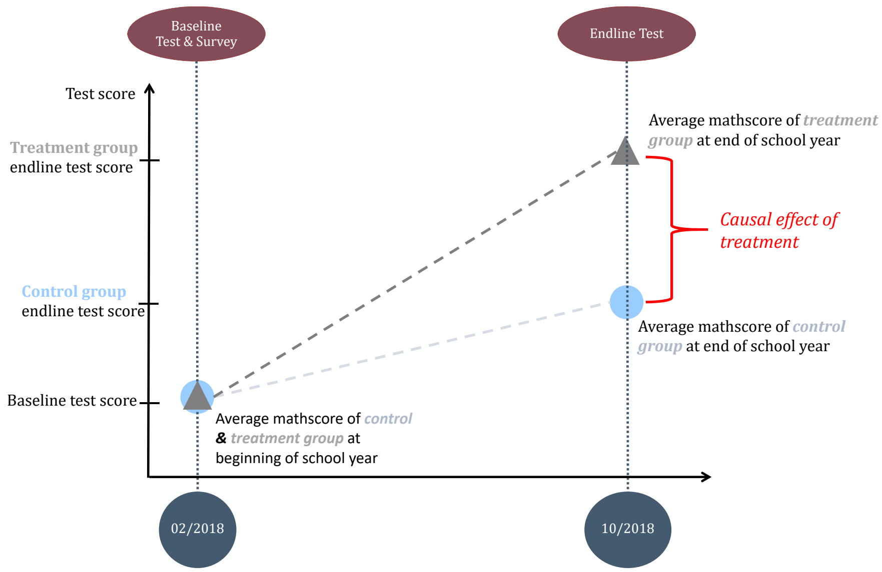

Figure 2: Identifying Causal Effects in a RCT

to the intervention, as illustrated in Figure 2. These differences reflect the true causal impact of the intervention. Yet, in non-experimental settings it is very challenging to find a valid counterfactual because any (intended or unintended) differences between treatment and control group would potentially confound the resulting estimate. The causal impact of the programme would be overestimated, for instance, if particularly well organized schools are selected as beneficiaries, whose students are then compared to students attending lower quality institutions. But the causal impact may also be underestimated if the NGO focuses on highly deprived school districts and compares them to an average school district with larger resources.

The cleanest solution to rule out any difference between treatment and control group is a random assignment mechanism. Random assignment guards against any intended or unintended selection and therefore ensures that the control and treatment group are on average identical. The evaluation of Consciente's CAL-project will be designed as a Randomized Controlled Trial (RCT), that is a methodological setup that explicitly builds on the random assignment of the intervention among potential recipients (e.g. Duflo et al., 2007).

While – by construction – a properly designed RCT ensures on average identical treatment and control groups, its validity depends on a second key assumption, the *Stable Unit Treatment Value Assumption (SUTVA)*. This assumption would be violated if the treatment causes spillover effects on the control group; a typical example would be medication which reduces the contagion risk among the control group (see Miguel and Kremer, 2004). Considering the nature of the intervention, we do not expect meaningful spillover effects on the control group in our setting (also see Section 5).

5 Sampling and Randomization

Starting point are all primary schools in Morazán, i.e. 300 schools. Consciente faces oversubscription, meaning that it cannot reach all eligible beneficiaries due to limited financial resources. Accordingly, the randomization schedule in this evaluation is based on oversubscription. From an ethical perspective, this has the favourable feature that it does not affect the number of beneficiaries reached by the NGO. We now briefly summarize the randomization schedule, which is illustrated in Figure 3.

Pre-selection. In a first step, we conduct a pre-selection of primary schools based on the following four criteria (ordered from most to least restrictive):

- ***School size:*** Schools with integrated classes (across grades) or gaps in their grade structure (i.e. not at least one class per grade) are excluded.
- ***Security:*** Schools located in areas dominated by criminal gangs are excluded.
- ***Accessibility:*** Schools inaccessible by car are excluded.
- ***Electricity:*** Schools without electricity are excluded.

After this pre-selection 57 schools, 320 classes and about 6400 students in grades 3–6 remain in our sample.

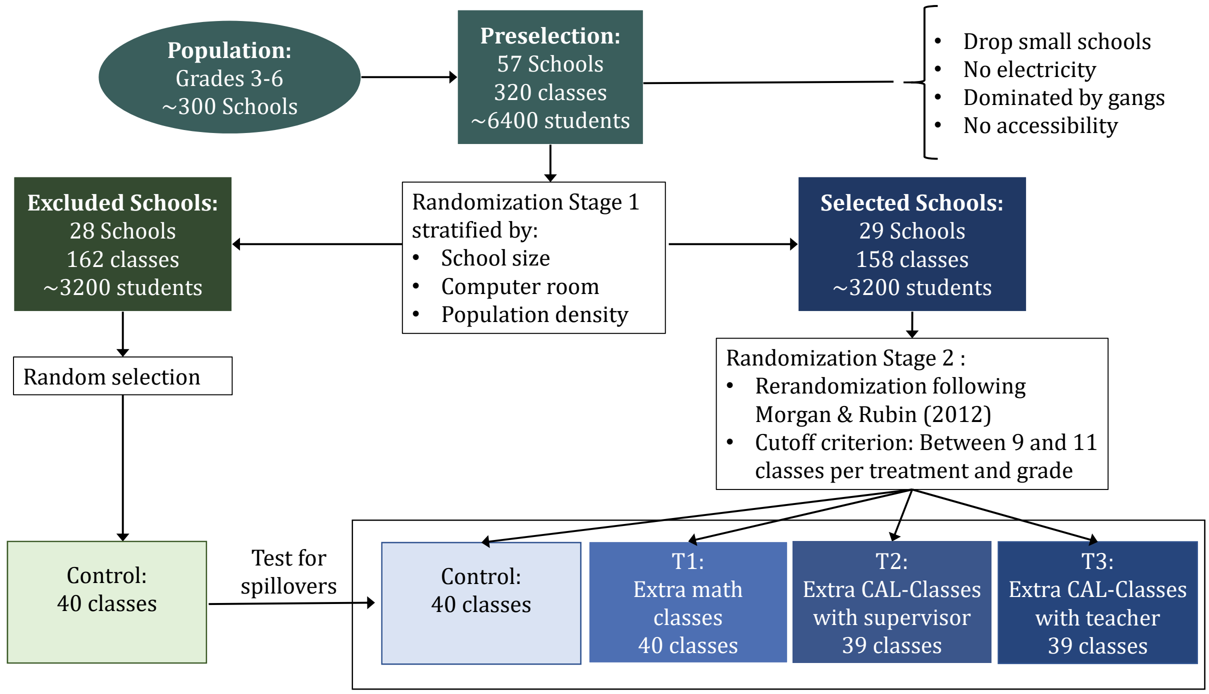

Figure 3: Our Sampling and Randomization Schedule

Randomization stage 1. Project resources do not allow to install computer rooms in these 57 schools. Therefore, 29 of these 57 schools are randomly chosen to be part of the study. We stratify by school size, population density and the existence of a computer room, to increase the precision of our estimates. Figure 4 maps the location of the 57 schools eligible for treatment; the coloured symbols mark schools that were selected in the first randomization stage, while white symbols mark schools that were excluded in the first randomization stage.

Randomization stage 2. In a second step, we randomize on the class level, i.e. 158 classes in the selected 29 schools are subject to the randomization. These classes are randomly assigned to the different treatment arms, with each treatment resp. control group comprising 39 or 40 classes. We use rerandomization (see Morgan and Rubin, 2012) to make sure that

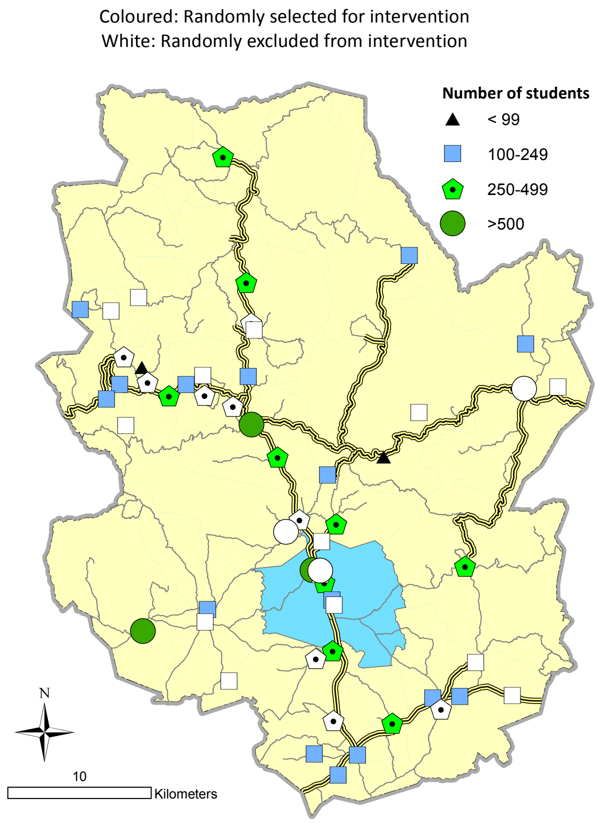

Figure 4: Selection of Schools in Morazán (in blue the capital San Francisco Gotera)

treatments are balanced across schools and grades. A balanced assignment across schools and grades has two advantages: An efficient use of the NGOs resources due to lower logistic costs, and higher statistical power when analysing effect heterogeneity across age groups. On the downside, randomization within schools increases the risk of spillovers between control group and treatment groups, which may bias the estimated effect. Although no evidence for such spillovers is reported in the related literature (e.g. Lai et al., 2015), we take different measures (e.g. close monitoring) to actively prevent spillovers. Furthermore, our setup allows to measure potential spillovers: We randomly choose 40 classes from the 28 schools that were excluded in the first randomization stage, and ask them to complete the baseline and endline exams. Comparing their results with the results of control classes in the "treatment" schools yields a quantitative estimate of potential spillovers.

6 Estimation Strategy

To quantify the causal effects of the different treatments, we estimate the following regression:

$$NS_{ics}^{NOV} = \beta_1 * T1_{cs} + \beta_2 * T2_{cs} + \beta_3 * T3_{cs} + \delta * NS_{ics}^{FEB} + \gamma * X_{ics}^{Feb} + \lambda_s + \epsilon_{ics}$$

- NS_{ics}^{NOV} represents numeracy skills of student i in class c and school s as measured at the end of the school year in November 2018; see Section 3 for further explanations.
- $T1_{cs}$, $T2_{cs}$ and $T3_{cs}$ indicate whether class c in school s received extra lessons with a teacher (T1), extra CAL-lessons with a supervisor (T2), or extra CAL-lessons with a teacher (T3).
- NS_{ics}^{FEB} is the test score of the baseline exam conducted in February 2018.
- X_{ics}^{Feb} is a vector of demographic and socio-economic characteristics. These are collected in a short questionnaire distributed prior to the baseline test in February.
- λ_s are a school dummies to absorb school specific characteristics.
- ϵ_{ics} is the error term.

β_1, β_2 and β_3 represent the causal effect of the corresponding treatment. If β_3 is significantly larger than β_1, this would suggest that the potential increase in learning outcomes is mainly attributable to the use of the software. The opposite would suggest that the use of the software is not important and that it is more effective to simply introduce refresher lessons in the afternoon without computers. Furthermore, we can address the question, whether the CAL-software and teacher skills are complements. This can be done by testing whether the ratio of the effects, i.e. β_3/β_2, is larger or smaller than the ratio of the respective costs of the two intervention arms, i.e. $cost(T3)/cost(T2)$. If the ratio β_3/β_2 is larger than $cost(T3)/cost(T2)$, teaching skills and software are complements and it is cost-effective to implement the treatment based on CAL-lessons and maths teachers. In the opposite case, software and teaching skills are not strongly complementary and it would be more cost-efficient to implement the treatment based on CAL-lessons and supervisors.

References

de Ayala, R. (2008). *The Theory and Practice of Item Response Theory*. Guilford Press, New York.

Duflo, E., Glennerster, R., and Kremer, M. (2007). Using Randomization in Development Economics Research: A Toolkit. In Schulz, P. and Strauss, J., editors, *Handbook of Development Economics*, volume 4, pages 3895–3962. North-Holland, Amsterdam.

Lai, F., Luo, R., Zhang, L., Huang, X., and Rozelle, S. (2015). Does Computer-Assisted Learning Improve Learning Outcomes? Evidence from a Randomized Experiment in Migrant Schools in Beijing. *Ecomomics of Education Review*, 47:37–48.

Miguel, E. and Kremer, M. (2004). Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities. *Econometrica*, 72(1):159–217.

MINED (2013). Elementos para el desarrollo del modelo pedagogico del sistema educativo nacional – escuela inclusiva de tiempo pleno. Ministerio de la Educacion de El Salvador, published online: `https://www.mined.gob.sv/jdownloads/Institucional/modelopedagogico.pdf` (last accessed: 14.01.2018).

Morgan, K. and Rubin, D. (2012). Rerandomization to Improve Covariate Balance in Experiments. *The Annals of Statistics*, 40(2):1263–1282.

OECD (2016). Skills in Ibero-America: Insights from PISA 2012. Organization for Economic Cooperation and Development, , published online: `https://www.oecd.org/education/Skills-IberoAmerica.pdf` (last accessed: 14.01.2018).

Snilstveit, B., Stevenson, J., Phillips, D., Vojtkova, M., Gallagher, E., Schmidt, T., Jobse, H., Geelen, M., Pastorello, M., and Eyers, J. (2015). "Interventions for Improving Learning Outcomes and Access to Education in Low- and Middle- Income Countries: A Systematic Review". 3ie Systematic Review 24. London: International Initiative for Impact Evaluation (3ie).

Are Teachers and Learning Software Complements or Substitutes? Evidence from a Randomized Experiment in El Salvador

Last registered on April 12, 2018

Pre-trial Fields

Trial Information

General Information

Title

Are Teachers and Learning Software Complements or Substitutes? Evidence from a Randomized Experiment in El Salvador

RCT ID

AEARCTR-0002789

Initial registration date

April 10, 2018

Last updated

April 12, 2018 6:04 PM EDT

Location(s)

Country

El Salvador

Region

Morazan

Primary Investigator

Name

Konstantin Buechel

Affiliation

University of Bern

Email

konstantin.buechel@vwi.unibe.ch

Other Primary Investigator(s)

PI Name

Daniel Steffen

PI Affiliation

University of Bern

PI Email

daniel.steffen@vwi.unibe.ch

PI Name

Christoph Kuehnhanss

PI Affiliation

University of Bern

PI Email

christoph.kuehnhanss@soz.unibe.ch

PI Name

Aymo Brunetti

PI Affiliation

University of Bern

PI Email

aymo.brunetti@vwi.unibe.ch

PI Name

Martina Jakob

PI Affiliation

University of Bern

PI Email

martina.jakob@soz.unibe.ch

Additional Trial Information

Status

On going

Start date

2018-02-12

End date

2018-10-01

Keywords

Education

Additional Keywords

Computer-assisted learning, numeracy skills

JEL code(s)

I20, I21, I28, O15

Secondary IDs

Abstract

In this study we experimentally evaluate an education program in the Salvadorean district Morazan that expands schooling by an additional 180 minutes per week. The program

comprises three different interventions that target primary school children in grades 3 to 6: (1) addtional math lessons taught by contract teachers, (2) additional math lessons taught by contract teachers using the software "Khan Academy", and (3) additional math lessons based on the software "Khan Academy" supervised by technical staff (but not teachers). This setup allows us to study the degree of complementarity/substitutability between computer-assisted learning software and pedagogically trained teachers. In particular, we appraise whether the (potential) impact on learning outcomes is primarily attributable to additional math classes or the use of computers, and further compare the cost-effectiveness of the three different versions of this school expansion program.

External Link(s)

Registration Citation

Citation

Brunetti, Aymo et al. 2018. "Are Teachers and Learning Software Complements or Substitutes? Evidence from a Randomized Experiment in El Salvador." AEA RCT Registry. April 12. https://www.socialscienceregistry.org/trials/2789/history/28149

Sponsors & Partners

Sponsor(s)

Sponsor name

NADEL - Center for Development and Cooperation, ETH Zurich

Sponsor location

Zurich, Switzerland

Sponsor Url

http://www.nadel.ethz.ch/

Sponsor name

Swiss Agency for Development and Cooperation

Sponsor location

Bern, Switzerland

Sponsor Url

https://www.eda.admin.ch/sdc

Partner(s)

Name

Consciente

Type

ngo

Url

https://consciente.ch/

Name

Ministerio de Educacion de El Salvador (MINED)

Type

government

Url

https://www.mined.gob.sv/

Experimental Details

Interventions

Intervention(s)

This schooling intervention, which is implemented by the NGO Consciente in collaboration with the regional Ministry of Education (MINED), comprises three different versions of additional math classes that complement regular classes of 3 to 6 graders (i.e. primary school level). The additional math classes cover two lessons of 90 minutes per week and are implemented for a period of 24 weeks (mid April-- end of September 2018) during the school year 2018.

The first treatment arm offers additional math classes conducted by contract teachers. The contract teachers are hired by Consciente and teach refresher courses that repeat the math curricula of lower grades.

The second and the third treatment are additional math classes based on the computer-assisted learning software "Khan Academy". Students work with the Spanish Lite-Offline version of the software, which allows them to learn independently and -- more important -- at their own level and pace. The difference between these two latter treatments is that the additional computer classes are either conducted by a temporarily contracted supervisor or by a temporarily contracted math teacher. Supervisors provide technical support but will not assist with questions regarding math. Teachers, in contrast, are also allowed to explain mathematical concepts to students, and hence complement the instructions/explanations provided by the software.

Intervention Start Date

2018-04-16

Intervention End Date

2018-10-01

Primary Outcomes

Primary Outcomes (end points)

Learning outcomes in math

Primary Outcomes (explanation)

Learning outcomes in math are measured in two consecutive "pencil & paper" tests that last 45 minutes each; the first part covers material taught in grades 1 to 3 (identical across all grades), while the second part is specifically designed for each grade and tests more advanced material. Both parts comprise 30 items that were selected from various sources including (a.) official text books of El Salvador, (b.) publicly available items from the STAR evaluations in California, (c.) publicly available items from the VERA evaluations in Germany, and (d.) exercises from the Swiss textbook MATHWELT. The items are selected such that they reflect the weighting in the official curriculum: 60--65% number sense and arithmetic, 30% geometry and measurement, 5--10% data and probability.

Secondary Outcomes

Secondary Outcomes (end points)

School attendance

Secondary Outcomes (explanation)

Data on school attendance for children and teachers is collected by monitoring staff who visit both regular classes and the additional treatment classes without prior notice. The attendance data is based on at least four surprise visits per class equally distributed across the months between April and September.

Experimental Design

Experimental Design

Starting point are all primary schools in Morazan, i.e. about 300 schools. Our partner NGO faces oversubscription, meaning that it cannot reach all eligible beneficiaries due to limited financial resources.

Pre-selection: In a first step, we pre-select primary schools based on the following four criteria (ordered from most to least restrictive):
- (i) School size: Schools with integrated classes (across grades) or gaps in their grade structure (i.e. not at least one class per grade) are excluded.
- (ii) Security: Schools located in areas dominated by criminal gangs are excluded.
- (iii) Accessibility: Schools inaccessible by car are excluded.
- (iv) Electricity: Schools without electricity are excluded.

After this pre-selection 57 schools, 317 classes, and about 6300 students in grades three to six remain in our sample.

Randomization stage 1: Project resources do not allow to install computer rooms in all 57 pre-selected schools. Therefore, 29 of these 57 schools are randomly chosen to be part of the study. We stratify by school size (3 strata, at least 1 / 2 / 3 classes per grade), population density (4 strata, quartiles) and the existence of a computer room (2 strata, yes / no).

Randomization stage 2: 158 classes in the 29 remaining schools are randomly assigned to the different treatment arms, with each treatment/control group comprising 39 or 40 classes. We use rerandomization (see Morgan and Rubin, 2012) to make sure that treatments are balanced across schools and grades. The threshold criterion is defined as follows: 1) Each school has to receive all three treatment arms and a control class, 2) each treatment/control arm comprises at least 9 classes and at most 11 classes per grade, with a total of 39 or 40 classes.

Randomization within schools increases the risk of spillovers between control group and treatment groups, which may bias the estimated effect. Although no evidence for such spillovers is reported in the literature on related interventions, we take different measures (e.g. close monitoring, extensive briefing of staff) to actively prevent spillovers. Furthermore, we randomly match 40 classes from the 28 "pure control" schools that were excluded in the first randomization stage to the control classes in the "treatment" schools. The children in these pure control classes are also asked to complete the baseline and endline exams.

Experimental Design Details

Randomization Method

Randomization done in the office using Stata (Version 14.0/SE).

Randomization Unit

School classes

Was the treatment clustered?

Yes

Experiment Characteristics

Sample size: planned number of clusters

158 classes in 29 schools, plus 40 classes in pure control schools.

Sample size: planned number of observations

Treatment 1: 40 classes and 750 children; Treatment 2: 39 classes and 787 children; Treatment 3: 39 classes and 740 children; Control (Treatment Schools): 40 classes and 738 children; Pure Control: 40 classes and 747 children

Sample size (or number of clusters) by treatment arms

Treatment 1 (2 x 90min per week with contract teacher): 40 classes
Treatment 2 (2 x 90min per week with CAL software and contract teacher): 39 classes
Treatment 3 (2 x 90min per week with CAL software and supervisor): 39 classes
Control classes (only regular math lessons in "treatment" schools): 40 classes
Pure control classes (only regular math lessons in "control" schools): 40 classes

Minimum detectable effect size for main outcomes (accounting for sample design and clustering)

MDE = 0.15--0.25 standard deviations. Calculations based on formula by Bloom (2007) and the following parameter values: power=80%; alpha (level of significance)=0.05; rho (intra-cluster correlation)=0.25; R2b (share of between-variance absorbed by baseline scores): 0.4--0.8 ; R2w (share of within-variance absorbed by baseline scores: 0.1--0.6; P (share of control clusters)=0.5; n (observations per cluster)=20

Supporting Documents and Materials

Documents

IRB

INSTITUTIONAL REVIEW BOARDS (IRBs)

IRB Name

IRB Approval Date

IRB Approval Number

Analysis Plan

Analysis Plan Documents

Post-trial Fields

Post-trial Information

Study Withdrawal

This trial has not been withdrawn.

Intervention

Is the intervention completed?

No

Is data collection complete?

Data Publication

Data Publication

Is public data available?

No

Is there a restricted access data set available on request?

Program Files

Program Files

Reports and Papers

Preliminary Reports

Relevant Papers