Inter-American Development Bank

IDB WORKING PAPER SERIES Nº IDB-WP-00958

Inquiry and Problem Based Pedagogy: Evidence from 10 Field Experiments

Rosangela Bando
Emma Näslund-Hadley
Paul Gertler

Inter-American Development Bank
Office of Strategic Planning and Development Effectiveness
Education Division

October, 2018

Inquiry and Problem Based Pedagogy: Evidence from 10 Field Experiments

Rosangela Bando
Emma Näslund-Hadley
Paul Gertler

October, 2018

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library
Bando, Rosangela.
Inquiry and problem based pedagogy: evidence from 10 field experiments / Rosangela Bando, Emma Näslund-Hadley, Paul J. Gertler.
p. cm. — (IDB Working Paper Series ; 958)
Includes bibliographic references.
1. Problem-based learning-Latin America. 2. Inquiry-based learning-Latin America. 3. Teacher effectiveness-Latin America. 4. Mathematics-Study and teaching (Elementary)-Latin America. 5. Science-Study and teaching (Elementary)-Latin America. I. Näslund-Hadley, Emma. II. Gertler, Paul, 1955- III. Inter-American Development Bank. Education Division. IV. Inter-American Development Bank. Office of Strategic Planning and Development Effectiveness. V. Title. VI. Series.
IDB-WP-958

http://www.iadb.org

1300 New York Ave NW, Washington, DC ZIP 20577

Rosangela Bando rosangelab@iadb.org

Inquiry and Problem Based Pedagogy: Evidence from 10 Field Experiments

Rosangela Bando

Emma Näslund-Hadley

Paul J. Gertler

October 1st, 2018

Abstract: We analyze evidence from 10 at-scale field experiments in four countries on the effect of inquiry- and problem-based pedagogy (IPP) on students' math and science test scores. IPP creates active problem solving opportunities in settings that derive meaning to the child. Students learn by collaboratively solving real life authentic problems, developing explanations, and communicating ideas. We find that IPP increased math test scores by 0.18 standard deviations and science test scores by 0.16 standard deviations after 7 months. Moreover, the results are robust across a wide set of geographic, socio-economic, and cultural, age/grade, and teacher background contexts.

Keywords: Education, pedagogy, teacher effectiveness, mathematics, science, elementary education, inquiry-based learning, problem-based learning, early education.

JEL Codes: C93, I25, I38, O15

Acknowledgments: We gratefully acknowledge the outstanding research assistance of Gabriel Englander. The authors are thankful to the Inter-American Development Bank and the Government of Japan for funding the research presented in this paper. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the Inter-American Development Bank, its Board of Directors, or the countries they represent. The authors have no conflicts of interests or financial or material interests in the results. Registry numbers: (i) Paraguay 2011 and 2013 IPA IRB Protocol #: 241.11May-007 and AEA RCT # [AEARCTR-0002947]; (ii) Peru 2012 Mathematics IPA IRB Protocol #: 12February-003 and 2014 IPA IRB Protocol #: 212.10April-002 for 2012 both with AEA RCT # [AEARCTR-0000365]; (iii) Peru Science 2014 Mathematics IPA IRB Protocol #: 12February-003 and AEA RCT#[AEARCTR-0000379]; (iv) Peru Science 2012 IPA IRB Protocol #: 215.10April-002 and AEA RCT#[AEARCTR-0002960]; (iv) Belize [ISCR/H/2/71] and AEA RCT#[AEARCTR-0002959]. The data for Argentina was provided to the authors by the government of Argentina, who executed the implementation. Thus, we do not have registry or IRB information.

Author Affiliations: Bando, Inter-American Development Bank, rosangelab@iadb.org; Näslund-Hadley, Inter-American Development Bank, EMMAN@iadb.org; Gertler, UC Berkeley and NBER, gertler@berkeley.edu.

1. Introduction

The education literature has long emphasized that students learn better when they play an active role in the learning process through doable tasks with social interaction (Vygotsky's 1978; Lowery, 1998; Zion and Selzak 2005; Furtak et al. 2012).[1] Meta-analyses confirm that traditional lecturing with passive listening is not conducive to fostering critical thinking, inspiring interest, or changing attitudes (Bligh 2000, Näsund-Hadley et al. 2014, and Bruns and Luque 2015). Rather, learning through activities, group work, and interactive class conversations is strongly associated with greater learning (Freeman et al., 2014).

One such active leaning approach is inquiry and problem-based pedagogy (IPP).[2] IPP creates active problem-solving opportunities in settings that derive meaning to the child. Students learn by collaboratively solving real life authentic problems, developing explanations, and communicating ideas (Hmelo-Silver et al 2007). They are taught to search for information from different sources, both text-based resources and from gathering their own data, and to develop problem-solving skills by collaboratively engaging in investigations. This approach helps solidify concepts through the child's exploration of research questions, the production and collection of evidence, the construction of theories based on evidence, and the development of explanations.

Teachers play critical roles in IPP. When done well, IPP includes important elements of explicit instruction and scaffolding (Hmelo-Silver, 2004; Edelson, 2001). Teachers facilitate learning by guiding students through a series of steps and explicitly relating learning to students' prior knowledge and experiences (Hmelo-Silver, 2004). They provide structure and scaffolding to help students do more complex activities by not only helping students to do the activity, but also helping them to comprehend why they are doing it and how it is related to the set of core concepts that the students are exploring (Vygotsky 1978). Teachers guide learners through complex tasks with explicit instructions on that are relevant to the problem at hand (Edelson 2001).

The difference between an IPP and a traditional lesson is illustrated by a unit on the skeletal system in the 4th grade in Argentina (IDB, 2018a). In traditional classrooms, students

[1] More generally, there is board consensus that a teacher's pedagogical style and the quality of teacher-student interactions are key inputs into student learning. See, for example, Angrist and Lavy (2001), Banerjee et al. (2011), Duflo, Dupas and Kremer (2012), Thomas and Staiger, (2012), Kremer, Brannen and Glennerster (2013), Yann, Cahuc and Shleifer, (2013), Murnane and Gaminian (2014) and Araujo et al. (2016).

[2] While Inquiry-Based and Problem-Based learning have different origins, they are in practice almost identical. Inquiry Learning has its roots in scientific research (Dostál, 2015), and Problem-Based Learning has its roots in medical education (Barrows and Tamblyn, 1980; Schmidt, 1983).

copy facts about bone tissues and the names of the 206 bones of the human skeleton that teachers have written on the blackboard into notebooks. They then answer questions about bones based on the lectures and material that they have read in textbooks and write those answers in a notebook or on a worksheet. Teachers then review their answers. In IPP classrooms, teachers pose research questions, and guide students through the formulation and testing of hypotheses to explore the questions. A motivating question might be what do bones help people do? Students then research facts about bones from texts and other sources from which they devise. A research question is what would happen if humans did not have bones? Students might answer this question by creating 3D clay figures and make predictions about how long they could stand with and without toothpick bones. Another question is how does loss of calcium affect bone strength. Here students soak chicken bones in vinegar for different lengths of time to extract different amounts of calcium, concluding that the more calcium a bone loses, the more it will bend.

In mathematics, the contrast between the IPP lessons and the traditional lessons is equally stark. For example, consider a lesson on ratios in the sixth grade in Belize (IDB, 2018b). In a traditional classroom, the lesson begins with a lecture that covers the definition of a ratio and how to solve simple mathematics problems involving ratios. The students then spent the rest of the class solving similar ratios. Upon completion, teachers review their work. In an IPP classroom, the teacher first uses examples to convey the concept of a ratio (e.g. the ratio of students with long and short sleeved shirts). Students then work in pairs to come up with definitions of a ratio. The teacher also provides them with a series of exercises to explore the use of ratios in everyday life. For example, pairs of students might be asked to investigate how many Cuisenaire rods of different colors are needed to measure the length of their desks and the relationships (shares) between the numbers of rods of different colors.[3] The small group exploration is followed by a teacher-led class discussion about the ratios of different color rods. The lesson ends with students revising their definitions of a ratio and a class conversation guided by the teacher to arrive on a joint definition and properties of ratios.

IPP is organized around core concepts that are developed over many lessons. This helps connect the topics in order to develop a deeper understanding that serves as a foundation for learning in subsequent academic years. For example, a core mathematics concept taught in early grades might be to develop an understanding of measurement, including how objects can be measured in standard and non-standard units and how the length of objects can be compared to

[3] Cuisenaire rods are a learning aid that help students visualize mathematics and learn mathematical concepts, including algebra, addition, subtraction, multiplication, fractions and proportional reasoning.

each other. In science, an example of a core concept is the interaction and dynamics of ecosystems, including the relationship among organisms within an ecosystem and how living things need one another to survive.

In this paper, we use data from 10 randomized field experiments in four Latin American countries to estimate the effect of IPP compared to traditional pedagogy that employs lecture with passive listening that emphasizes memorization on preschool and primary school student learning in math and science.[4] This approach not only provides strong causal evidence, but also an unusual degree of external validity. One of the challenges faced when evaluating specific programs is the applicability of the evidence to other contexts (Fisher, 1935, Campbell, 1957, Pritchett and Sandefur, 2013, Manski, 2013, Athey and Imbens, 2017).[5] These 10 experiments allow us to examine the effects of IPP across a wide set of geographic, socio-economic, teacher background, and age/grade contexts (i.e. preschool, 3rd grade and 4th grade).

2. The Interventions

This study focuses on 10 IPP randomized field experiments in four Latin America countries: Argentina, Belize, Paraguay, and Peru.[6] These countries represent a large range in income (GDP per capita), from US$4,078 in Paraguay to US$12,440 in Argentina (World Bank, 2018), and a large range in population, from 366,954 in Belize to 43,847,430 in Argentina (World Bank, 2018). Like many countries, these four face challenges with education quality. Their PISA scores show learning severe deficits when compared to OECD countries (Bos, 2016a)[7].

Although the specifics of the interventions varied by grade level, national curriculum, and subject area (mathematics or science), all shared three central elements of IPP: (1) instruction

[4] One of the eight programs that we study in this paper was previously investigated in Beuermann et al. (2013) who found that IPP led to a 0.18 standard deviation increase test scores among 3rd graders in Peru.
[5] A current active area of research is the identification of relevant dimensions and relationships to extrapolate lessons learned. For example, Dehejia, Pop-Eleches y Samii (2015), Hunt (2015) and Vival (2014) study extrapolation error. Examples of other studies that include multiple sites to increase external validity external validity include Cruces and Galiani (2007), Banerjee, Karlan and Zinman (2015), Dehejia, Pop-Eleches y Samii (2015), and Galiani et al. (2017).
[6] Appendix Table 1 describes the details of each the interventions including year, target population and grade, didactic materials, hours of teacher training, and frequency of supervision.
[7] The Programme for International Student Assessment (PISA) is a triennial international survey by the Organisation for Economic Co-operation and Development (OECD). It aims to evaluate education systems worldwide by testing 15-year-old students. Its main goal is to provide data comparable across countries to improve policy. It tests mathematics, science, and reading. PISA emphasizes testing on what students can do with what they know, rather than recall of facts. Thus, it measures problem solving and cognition in daily life. For more information please see http://www.oecd.org/pisa/.

was organized around core concepts that were developed over many lessons, (2) classes organized around inquiry and problem solving opportunities, and (3) use of structure and scaffolding to help student do more complex activities and make sure that they have close guidance.

All of the programs were implemented at the class level, except for Peru 2014. In Peru 2014 individual tutors were used, making the program much more expensive. Each program trained teachers (tutors) in IPP methods and lesson plans, provided didactic materials to enhance learning through hands-on activities, and provided ongoing supervision. All programs included a detailed lesson by lesson curriculum and a minimum of 20 hours of teacher professional development. The sessions aimed to have teachers learn through practice and interactions with other teachers. In addition, all programs provided continuous in-school teacher support.

3. Experimental Designs

Although the details of each study differ, all studies employed a cluster (school level) randomized design, except for Peru 2014. Peru 2014 randomized students at the individual level.[8] Study schools in Argentina and Peru were randomly selected from the respective country-year universe of schools with students enrolled in the grade of interest. In Paraguay and Belize study schools were selected from the universe of eligible schools that had students in the grade of interest and that additionally volunteered to participate. Schools were compliant with treatment assignment in all cases except for one control school in Paraguay 2011 where teachers received training. For that case, we present intention to treat estimates.

Except for Peru 2014, all students in the target grades in the study schools participated in the study. Peru 2014 instead enrolled students that performed in the bottom half of the test score distribution.

All studies, except for Belize 2015, collected panel data at the student level with one survey before treatment, and another after treatment. In all studies the same group was surveyed before and after the intervention, except for Belize 2015 where baseline and follow up surveys were administered to different cohorts. The length of exposure to IPP was 7 months in all cases.

[8] Appendix Table 2 provides the details of each experimental design including sample frame, sample size in terms of number of schools and number of students, stratifications for random assignment, and timing of data collection.

The key outcome of interest is students' standardized test results. Each test was designed to measure the ability of students to understand and apply key mathematical and scientific concepts. Tests were adapted for each grade level and administered by an external evaluator, rather than the local teachers. Surveys of parents provided additional information about the student and family. Teacher and school level information was merged into the student level data base. Appendix Table A3 provides the definition for each variable used in the analysis.

4. Estimation Methods

We estimate the following regression specification for each country-year subject intervention:

$$y_{is} = \mu_s + \beta T_{is} + \varepsilon_{is} \qquad (1)$$

Where y_{is} denotes score for student *i* in strata *s*, μ_s is a strata fixed effect and ε_{is} is an error term. The variable T_{is} equals 1 if the students receives treatment and 0 otherwise. β represents the average difference in student scores between treatment and control units.

In order to improve precision and to test for robustness we estimate the treatment effects a number of different ways. We estimate a second version of (1) that includes lagged (baseline) individual test scores,

$$y_{ist} = \mu_s + \beta T_{is} + \gamma y_{ist-1} + \varepsilon_{is} \quad , \qquad (2)$$

in order to reduce residual variance and improve statistical power (Imbens and Wooldridge, 2009). We then estimate a third version by adding the controls listed in Appendix Table 2 for each experiment-year to model (2). Finally, we estimate a fourth model that instead uses student fixed effects,

$$y_{ist} = \mu_i + \beta T_{is} + \varepsilon_{is} \quad . \qquad (3)$$

For inference, we cluster errors at the school level. We estimate confidence intervals with a bootstrap approach for the Argentina Mathematics experiment to avoid over-rejection (Cameron and Miller 2015, Bertrand, Duflo, and Mullainathan 2004).

We take two approaches to providing estimates of the overall effects across all of the study samples; (i) regressions that pool all of the data from the country-year studies and (ii) a meta-analysis approach. The pooled model approach re-estimates the regression models presented above including strata dummies to vary by country and year. Thus, the pooled model estimates a common β across all strata, all countries and all years. We test for whether the data can be pooled into single model by the hypothesis that the β are equal for all samples.

Under the meta-analysis approach, we construct a weighted average of the individual country-year estimates weighted by the inverse of the variance of estimate (Sterne 2009). We test for cross-study heterogeneity using an I^2 statistic, which indicates the percentage of variation attributable to heterogeneity across studies (Higgins et al. 2003). I^2 takes values between 0 and 1, with 1 indicating high heterogeneity across studies.[9]

5. Baseline Balance and Sample Attrition

Descriptive statistics at baseline prior to the interventions and p-values for tests of the hypotheses that the means of the treatment group are equal to those of the control group show that the treatment and control groups are well balanced for all the study samples (Appendix Table A4). Mean math and sciences test scores in the treatment are not statistically different from the control group for all countries and years. Similarly, there are no differences for student age, whether bilingual, family assets, teacher's age and gender. There are gender imbalances in 3 of the 10 studies and in class size in 4 of the 10 studies.

The attrition rates by treatment and control groups, for each country except for Belize where we do not have a panel of students, show little evidence of selective attrition bias (Appendix Table A5). Student attrition over the 7-month period ranges from 3 percent in Paraguay 2011 to 17 percent in Argentina 2009. There is no differential attrition between treatment and control groups for all study samples except for Argentina 2009 where there was 4 percentage points more attrition in the control group than in the treatment group. Despite this, there appears to be no differences in the means of baseline test scores between treatment and control groups for the evaluation sample, i.e. the sample was found at endline (Appendix Table A4). Overall, we can reject only 5 of the 64 tests of the equality of treatment and control means at the 0.10 level.

[9] More specifically, $I^2=100\%*(Q-df)/Q$, where *Q* is the across study variation of impacts, and *df=k-1* denotes the degrees of freedom.

6. Results

The estimated effects of 7 months of IPP show meaningful positive impacts on both math and science test scores (Table 1). Moreover, the point estimates are robust to the specific estimation method used and our estimates become more precise in the models that add covariates (rows 2 and 3) and in the fixed effects models (row 4). Our preferred method that controls for strata, lagged test scores and other controls (row 3) generates estimates for math ranging from 0.10 in Argentina 2009 to 0.25 standard deviations in Belize 2015. Figure 1 depicts all of the country year point estimates and their 95 percent confidence intervals.

We take two approaches to providing estimates of the average impacts across all of the study samples based on our preferred specification using our preferred model in row 3 of Table 1. The pooled model estimates a 0.18 standard deviation impact on math test scores, a 0.16 standard deviation impact on science test scores, and a 0.17 standard deviation overall impact (Table 2). Moreover, we cannot reject the hypotheses that the estimated treatment effects are equal across all study samples for math (p=0.865), for science (p=0.582) and overall (p=0.867). The meta-analysis approach yields similar results. We estimate a 0.17 standard deviation impact on math test scores, 0.14 standard deviation on science test scores and a 0.16 overall impact (Table 2). The corresponding I^2 is 0 percent indicating rejection of cross-study heterogeneity. Figure 1 also depicts pooled estimates and their 95 percent confidence intervals.

Boys appear to benefit somewhat more than girls from IPP (Table 3). Specifically, the estimated gains in tests are about a quarter more for boys than girls in both math and science. In math, boys benefit more than girls, which is largely driven by Argentina 2009 and Paraguay 2011. In science, boys also benefit more than girls largely driven by Argentina 2009, Peru 2010 and Peru 2014.

Finally, we provide estimates of cost-effectiveness of each of the interventions (Table 1) and overall (Table 2).[10] Specifically, the cost of a 0.10 standard deviation increase in math test scores overall is $14.53 per student. This estimate varies from $6.90 per student in Argentina 2009 to $22.48 in Paraguay 2013. The cost of a 0.10 standard deviation increase in science test

[10] We use administrative data for each program to estimate incremental costs. Costs were in current dollars of the year the program was implemented. We use each countries Urban Consumer Price Index to normalize the costs as of March 2017. Costs are coverted in US dollars using the appropriate exchange rate. We include teacher training, didactic materials, and supervision costs. Training and material costs are depreciated over a 3-year period using straight-line depreciation.

scores overall is $14.64 per students and varies from $8.40 per student in Belize 2015 to $17.52 in Peru 2010.[11]

7. Discussion

We analyzed data from 10 field experiments in four countries to assess if teacher training to designed to change pedagogical practices from teacher centered lecturing with passive listening to student centered IPP learning processes improved student test scores. Our results strongly support the conclusion that implementing IPP learning in Latin America if not worldwide would greater enhance student learning in math and science.

We found that IPP resulted in a 0.18 standard deviation increase in math test scores overall and a 0.16 standard deviation increase in science test scores overall. Moreover, the effect sizes were not different in order of magnitude nor statistically statistical significance across these settings, suggesting a greater degree of external validity than most studies. This is important because programs also varied in terms of setting, intensity, the provision of complementary materials and teacher support. These results were present across two subject areas (math and science), 3 different grade levels (preschool, 3rd grade and 4th grade), and across four different countries. They were implemented under different educational systems and with teachers of differing backgrounds. For example, teacher experience ranged from average 11 years in Paraguay to 20 years in Peru. In Belize, 34 percent of teachers were male compared to no male teachers in the Peru mathematics program. The 2014 science program in Peru shows effects were also present when implemented as a tutoring program outside of the classroom. Further, the programs targeted students in different sociocultural conditions. For example, about 43 percent of children were bilingual in Paraguay contrasted with a large majority of monolingual children in Belize. In Paraguay, about 30 percent of children had mothers who had completed secondary education, compared 57 percent in Argentina.

These results are likely to be lower bound estimates of the effect of IPP on learning. First, this was the first time any of the teachers implemented IPP. One would expect teachers to get better over time. Second, the exposure period is only 7 months. One would expect that the more years that students are taught using IPP the larger the cumulative effects. Third, there are likely

[11] We exclude Peru 2014 from the overall cost-effectiveness analysis and discussion as it is a individual tutoring program and very costly relative to the other interventions which are teacher training with didactic materials.

dynamic complementarities in that students will be better prepared in subsequent grades and therefore learn more.

Our results are broadly consistent with the previous IPP literature. Qualitative assessments of the programs studied in this paper found that classes were more interactive and students were more involved in academic activities in treatment schools than students in control schools (UNESCO and Universidad Catolica 2010; Benson 2014; IPA, IDB and MINED 2014a and 2014b). Our findings are also in line with a broader set of studies in the education literature that suggest that some degree of inquiry-based classroom practices enhances learning (Lowery, 1998; Zion and Selzak 2005) and that guided inquiry is more effective than minimally guided instructional approaches (Furtak et al. 2012).

Finally, our results are also consistent with studies that assessed individualized instruction more generally as a pedagogical approach. Angrist and Lavy (2001) find that a teacher training program that aimed to promote a student-centered pedagogical approach led to an increase of 0.25 standard deviations in test scores after one year among fourth graders in secular schools but had no effect in religious schools in Jerusalem. Banerjee et al. (2007) found that by substituting two hours of class lecture per school day with individualized instruction provided by tutors led to improvements of 0.14 standard deviations after one academic year among first grade students in India. Duflo, Dupas and Kremer (2012) found that tracking students by ability increased learning by 0.16 standard deviations after 18 months among 1st grade students in Kenya.[12] McEwan (2015) estimates that interventions on the supply side that include a teacher training component improve teacher effectiveness by 0.12 standard deviations.

[12] However, tracking goes beyond tailored instruction because tracking involves peer effects.

References

Angrist, Joshua D. and Victor Lavy. 2001. “Does Teacher Training Affect Pupil Learning? Evidence from Matched Comparisons in Jerusalem Public Schools.” *Journal of Labor Economics*, 19 (2) 343-369.

Araujo, M. Caridad, Pedro Carneiro, Yyannú Cruz-Aguayo and Norbert Schady. 2016. ¨Teacher Quality and Learning Outcomes in Kindergarten.” *The Quarterly Journal of Economics* (2016), 1415-1453.

Athey, Susan and Imbens, Guido W. 2017. “The Econometrics of Randomized Experiments.” *Handbook of Field Experiments*, Volume 1 1st Edition Editors: Esther Duflo, Abhijit Banerjee

Banerjee, Abhijit, Dean Karlan, and Jonathan Zinman. 2015. "Six Randomized Evaluations of Microcredit: Introduction and Further Steps." *American Economic Journal: Applied Economics*, 7(1): 1-21.

Banerjee, Abhijit, Shawn Cole, Esther Duflo, Leigh Linden. 2007. “Remedying Education: Evidence from Two Randomized Experiments in India.” *The Quarterly Journal of Economics*, 122(3) 1235-1264.

Barrows, Howard S. and Robyn M. Tamblyn. 1980. Problem-based Learning. Springer Press: New York.

Benson, Nicole. 2014. “Tikichuela: Matemáticas en Mi Escuela Documento Anexo Componente Cualitativo 2014.” Innovations for Poverty Action. Asunción, Paraguay. Unpublished IDB Report.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. “How Much Should We Trust Differences-in-Differences Estimates?” *The Quarterly Journal of Economics* 119 (1): pp. 249–275.

Beuermann, Diether W., Emma Naslund-Hadley, Inder J. Ruprah and Jennelle Thompson. 2013. “The Pedagogy of Science and Environment: Experimental Evidence from Peru.” *The Journal of Development Studies,* DOI:10.1080/00220388.2012.754432

Bligh, Donald A. 2000. *What’s the Use of Lectures?* Jossey-Bass, San Francisco.

Bos, María Soledad, Alison Elías, Emiliana Vegas and Pablo Zoido. 2016a. ¨PISA Latin America and the Caribbean. How Much did the Region Improve?” Brief 2. Inter-American Development Bank. doi: 10.18235/0000543

Bruns, Barbara and Javier Luque. 2015. *How to Raise Student Learning in Latin America and the Caribbean.* doi:10.1596/978-1-4648-0151-8. Washington, DC: World Bank. License: Creative Commons Attribution CC BY 3.0 IGO

Cameron, Colin A and Doublas L. Miller. 2015. “A Practitioner’s Guide to Cluster-Robust Inference.” *Journal of Human Resources*, Vol.50, No. 2, pp.317-373.

Campbell, D. T. 1957. “Factors relevant for the validity of experiments in social settings”, *Psychological Bulletin* 54, pp. 297-312.

Cruces, Guillermo and Galiani, Sebastian. 2007. "Fertility and female labor supply in Latin America: New causal evidence," *Labour Economics*, Elsevier, vol. 14(3), pages 565-573, June.

Dehejia, Rajeev, Cristian Pop-Eleches and Cyrus Samii. 2015. “From Local to Global: External Validity in a Fertility Natural Experiment.” NBER Working Paper No. 21459.

Dostál, Jirí. 2015. *Inquiry-based instruction : Concept, essence, importance and contribution.* Olomouc: Palacký University.

Duflo, Esther; Pascaline Dupas and Michael Kremer. 2011. "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence form a Randomized Evaluation in Kenya." *American Economic Review* 101: 1739-1774.

Edelson, D. C. 2001. "Learning-for-use: A framework for integrating content and process learning in the design of inquiry activities*." Journal of Research in Science Teaching*, 38, 355–385.

Fisher, R. 1935. The Designs of Experiments, Oliver and Boyd, London.

Freeman, Scott, Sarah L. Eddy, Miles McDonough, Michelle K. Smith, Nnadozie Okoroafor, Hannah Jordt, and Mary Pat Wenderoth. 2014. "Active Learning Increases Student Performance in Science, Engineering, and Mathematics" *Proceedings of the National Academy of Sciences USA*, 111(23): 8410-8415.

Furtak, E., Seidel, T., Iverson, H., and Briggs, D. 2012. "Experimental and quasi-experimental studies of inquiry-based science teaching: A meta-analysis." *Review of Educational Research*, 82 (3), 300–329.

Galiani, Sebastian; Gertler, Paul; Undurraga, Raimundo; Cooper, Ryan; Martinez, Sebastian; Ross, Adam and. 2017. "Shelter from the Storm: Upgrading housing Infrastructure in Latin American Slums." Journal of Urban Economics, V. 98, 187-213. **Harris, D.N., and T.R. Sass.** 2011. "Teacher Training, Teacher Quality and Student Achievement." *Journal of Public Economics* 95:798–812.

Higgins, J. P. T., S. G. Thompson, J. J. Deeks, and D. G. Altman. 2003. "Measuring inconsistency in meta-analyses." *British Medical Journal* 327: 557–560.

Hmelo-Silver, Cindy E. 2004. "Problem-Based learning: What and how do students learn?" *Educational Psychology Review*. 16: 235-266.

Hmelo-Silver, Cindy E., Ravit Golan Duncan, and Clark A. Chinn. 2007. "Scaffolding and Achievement in Problem-based and Inquiry-Learning: A Response to Kirschner, Sweller, and Clark (2006)" Educational Psychologist 43(2): 99-107.

Hunt Allcott. 2015. "Site Selection Bias in Program Evaluation." *Quarterly Journal of Economics* 130 (3): 1117-1165.

Imbens, W. Guido and Jeffrey M. Wooldridge. 2009. "Recent Developments in the Econometrics of Program Evaluation." *Journal of Economic Literature* 47(1) 5-86.

Inter-American Development Bank (IDB). "AR-T1047 : Improvement of Natural Science and Mathematics Education" accessed March 14th, 2018a. https://www.iadb.org/en/project/AR-T1047

Inter-American Development Bank (IDB). "BL-L1018 : Education Quality Improvement" accessed March 14th, 2018b. https://www.iadb.org/en/project/BL-L101

Inter-American Development Bank (IDB). "PR-T1092 : Early Calculators" accessed March 14th, 2018c. https://www.iadb.org/en/project/PR-T1092

Inter-American Development Bank (IDB). "PR-T1095: Early Calculators" accessed March 14th, 2018d. https://www.iadb.org/en/project/PR-T1095

Inter-American Development Bank (IDB). "PR-T1128 : Mathematics in My School" accessed March 14th, 2018e. https://www.iadb.org/en/project/PR-T1128

Inter-American Development Bank (IDB). "PE-T1232 : Intercultural Preschool Math Iniciative" accessed March 14th, 2018f. https://www.iadb.org/en/project/PE-T1232

Inter-American Development Bank (IDB). "PE-T1243 : Improving the Teaching of Science and Environment Area" accessed March 14th, 2018g. https://www.iadb.org/en/project/PE-T1243

IPA, IDB and MINED. 2014a. "Análisis cualitativo de la evaluación experimental del programa MIMATE y del programa de mejora de la Educación de Ciencias II." Informe Final de Evaluación. Innovations for Poverty Action, Inter-American Development Bank, and Ministry of Education, Peru. Unpublished IDB Consultancy Report.

Lowery, L. 1998. *The biological basis of thinking and learning.* Berkeley: University of California.

Kremer, Michael, Conner Brannen, and Rachel Glennerster. 2013. "The Challenge of Education and Learning in the Developing World." *Science* 340: 297

Manski, C. 2013, *Public Policy in an Uncertain World*, Harvard University Press, Cambridge.

McEwan, Patrick J.2015. "Improving Learning in Primary Schools of Developing Countries. A Meta-Análisis of Randomized Experiments." *Review of Educational Research* 85(3) 353-394.

Murnane, Richard J. and Alejandro J. Ganimian. 2014. "Improving Educational Outcomes in Developing Countries: Lessons from Rigorous Evaluations." National Bureau of Economic Research Working Paper 20284.

Näslund-Hadley, Emma, Armando Loera Valera and Katelyn Ann Hepworth. 2014. "What Goes on Inside Latin American Math and Science Classrooms: A Video Study of Teaching Practices." *Global Education Review* 1(3):110-128.

Pritchett, Lant, and Justin Sandefur. 2013. "Context Matters for Size: Why External Validity Claims and Development Practice Don't Mix." Center for Global Development Working Paper 336.

Sterne, Jonathan A. C. Editor. 2009. *Meta-Analysis in Stata: An Updated Collection from the Stata Journal.* ISBN 978-1-59718-049-8. Stata Press. College Station, TX.

Thomas, Kane, and Douglas Staiger. 2012. ''Volatility in School Test Scores: Implications for Test-Based Accountability Systems.'' *Brookings Papers on Education Policy*, 235–283.

UNESCO and Universidad Catolica. 2010. "Programa de Mejora de la Enseñanza de las Ciencias Naturales y la Matemática." Volumes II, III and IV.

Vivalt, Eva. 2014. "How Much Can We Generalize from Impact Evaluation Results?", manuscript, New York University.

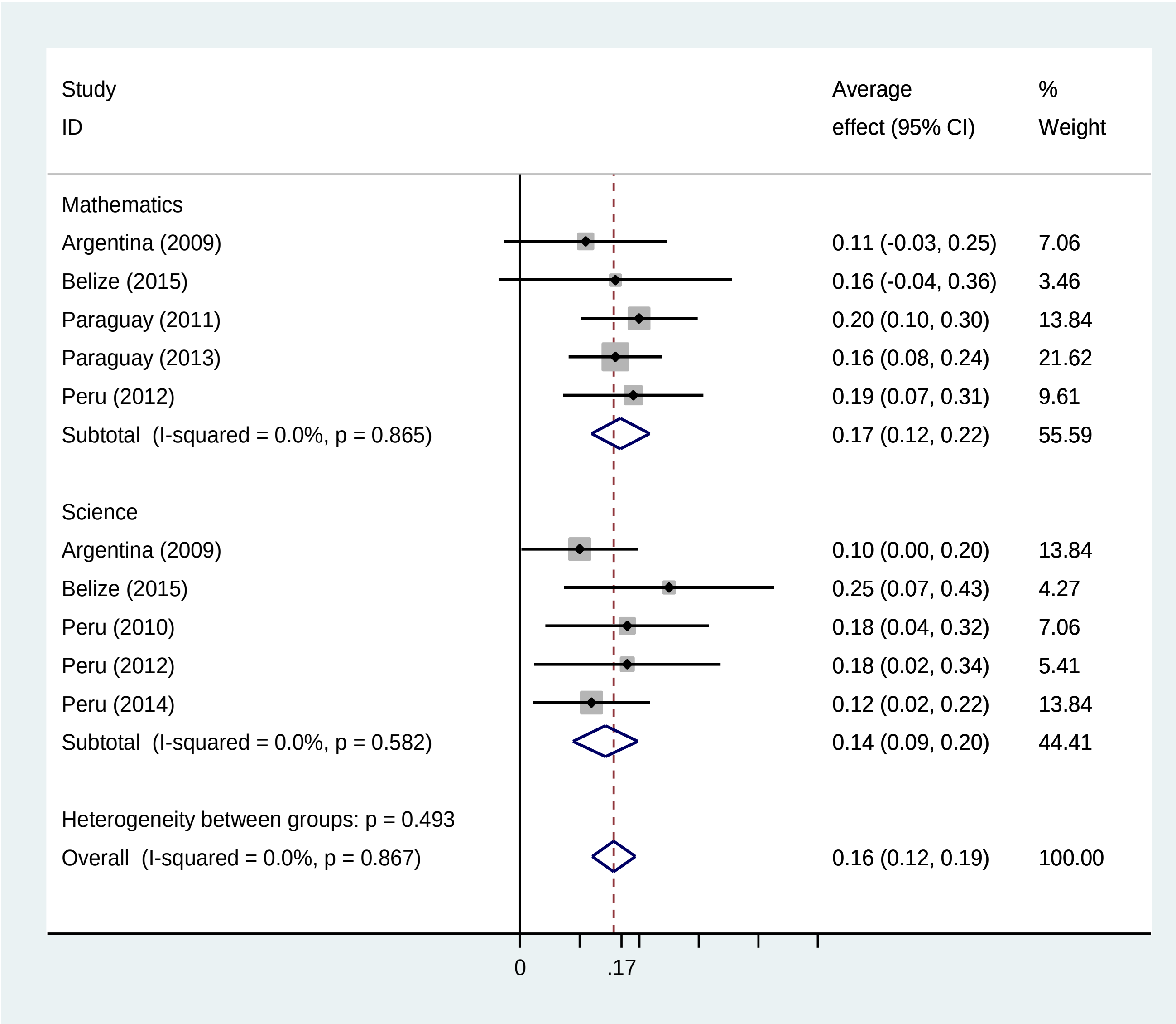
Vygotsky, L. S. 1978. Mind in society: The development of higher mental processes. Cambridge, MA: Harvard University Press.

World Bank. 2018. World Bank Open Data. World Development Indicators. https://data.worldbank.org [retrieved March 4th, 2018].

Yann, Algan, Pierre Cahuc, and Andrei Shleifer. 2013. "Teaching Practices and Social Capital." *American Economic Journal: Applied Economics,* 5(3): 189-210.

Zion and Selzak, 2005. "It takes two to tango: In dynamic inquiry, the self-directed student acts in association with the facilitating teacher." *Teaching and Teacher Education,* 21(7), 875-894.

Figure 1. Impact of IPP on student performance by country/year and overall.



Notes: This figure presents the estimated impact and 95% confidence regions IPP on math and science tests in standard deviations for each/country year and overall. The estimates correspond to those in row 3 of tables 1 and 2.

Table1. Estimated Country Treatment Effects (Impact on Student Test Scores) and Cost-Effectiveness.

	Mathematics					Science				
	Argentina 2009	**Belize 2015***	**Paraguay 2011**	**Paraguay 2013**	**Peru 2012**	**Argentina 2009**	**Belize 2015***	**Peru 2010**	**Peru 2012**	**Peru 2014**
Panel A: Estimated Treatment Effects										
1. Controls for Stratum	0.09 (0.06) [0.125]	0.13 (0.1) [0.188]	0.20 (0.07) [0.003]	0.12 (0.06) [0.029]	0.15 (0.08) [0.071]	0.08 (0.04) [0.056]	0.29 (0.1) [0.004]	0.21 (0.1) [0.043]	0.14 (0.09) [0.104]	0.11 (0.06) [0.049]
2. Adds lagged test scores to (1)	0.13 (0.06) [0.035]	0.16 (0.09) [0.071]	0.20 (0.05) [0.000]	0.17 (0.04) [0.000]	0.19 (0.06) [0.003]	0.08 (0.04) [0.054]	0.29 (0.09) [0.002]	0.17 (0.08) [0.032]	0.14 (0.08) [0,064]	0.12 (0.05) [0.024]
3. Adds additional Controls to (2)	0.11 (0.07) [0.012]	0.16 (0.1) [0.114]	0.20 (0.05) [0.000]	0.16 (0.04) [0.000]	0.19 (0.06) [0.004]	0.10 (0.05) [0.032]	0.25 (0.09) [0.007]	0.18 (0.07) [0.013]	0.18 (0.08) [0.022]	0.12 (0.05) [0.023]
4. Student Fixed Effects	0.20 (0.06) [0.000]		0.20 (0.04) [0.000]	0.19 (0.04) [0.000]	0.23 (0.05) [0.000]	0.09 (0.04) [0.024]		0.13 (0.05) [0.010]	0.14 (0.06) [0.018]	0.15 (0.08) [0.061]
Sample Size										
Number of Students	**1126**	**4457**	**2905**	**2888**	**2400**	**1927**	**4457**	**2392**	**2401**	**1127**
Number of Schools	**28**	**252**	**265**	**261**	**104**	**42**	**252**	**106**	**104**	**48**
Panel B. Cost-Effectiveness										
$/student for a 0.10 standard deviation increase in test scores	$6.90	$11.75	$17.58	$22.48	$19.68	$9.61	$8.40	$17.52	$17.20	$49.96

Notes: Each cell represents the estimated impact, standard error and *p*-value for the hypothesis that the coefficient equals zero from a different regression. The standard errors and *p*-values are clustered by school. The table shows standard errors in parenthesis and the *p*-values in brackets.

Table 2. Overall Estimated Treatment Effects (Impact on Student Test Scores) and Cost-Effectiveness.

	Mathematics		**Science**		**Combined**	
	Pooled	Meta	Pooled	Meta	Pooled	Meta
Panel A: Estimated Treatment Effects						
1. Controls for Stratum	0.14 (0.03) [0.000]	0.13 (0.03) [0.000]	0.16 (0.04) [0.000]	0.12 (0.03) [0.000]	0.15 (0.03) [0.000]	0.13 (0.02) [0.000]
2. Adds lagged test scores to (1)	0.18 (0.03) [0.000]	0.17 (0.02) [0.000]	0.16 (0.04) [0.000]	0.13 (0.03) [0.000]	0.17 (0.02) [0.000]	0.15 (0.02) [0.000]
3. Adds controls to (2)	0.18 (0.03) [0.000]	0.17 (0.03) [0.000]	0.16 (0.04) [0.000]	0.14 (0.03) [0.000]	0.17 (0.02) [0.000]	0.16 (0.02) [0.000]
4. Student Fixed Effects	0.21 (0.02) [0.000]	0.20 (0.02) [0.000]	0.13 (0.03) [0.000]	0.12 (0.03) [0.000]	0.17 (0.02) [0.000]	0.17 (0.02) [0.000]
Panel B: Cost Effectiveness						
Cost/Student for a 0.10 Standard Deviation Increase in Test Scores	$14.53		$14.64		$15.39	

Notes: Each cell represents the estimated impact, standard error and *p*-value for the hypothesis that the coefficient equals zero from a different regression. The standard errors and *p*-values are clustered by school. The table shows standard errors in parenthesis and the *p*-values in brackets.

Table 3. Estimated Country Treatment Effects (Impact on Student Test Scores) by Gender

Mathematics	**Argentina 2009**	**Belize 2015***	**Paraguay 2011**	**Paraguay 2013**	**Peru 2012**	**All Pooled**	**All Meta**
Boys	0.21 (0.1) [0.031]	0.12 (0.12) [0.341]	0.27 (0.06) [0.000]	0.15 (0.04) [0.001]	0.19 (0.07) [0.007]	0.20 (0.03) [0.000]	0.19 (0.03) [0.000]
Girls	0.01 (0.1) [0.929]	0.20 (0.12) [0.092]	0.13 (0.05) [0.017]	0.16 (0.06) [0.005]	0.18 (0.07) [0.009]	0.16 (0.03) [0.000]	0.18 (0.03) [0.000]
p-value Boys=Girls	0.087	0.584	0.020	0.724	0.597	0.000	
Science	**Argentina 2009**	**Belize 2015***	**Peru 2010**	**Peru 2012**	**Peru 2014**	**All Pooled**	**All Meta**
Boys	0.13 (0.04) [0.002]	0.27 (0.12) [0.019]	0.27 (0.08) [0.001]	0.18 (0.09) [0.029]	0.21 (0.07) [0.003]	0.20 (0.04) [0.003]	0.14 (0.03) [0.000]
Girls	0.07 (0.08) [0.408]	0.22 (0.1) [0.030]	0.10 (0.08) [0.252]	0.18 (0.08) [0.048]	0.01 (0.09) [0.870]	0.13 (0.04) [0.003]	0.11 (0.04) [0.003]
p-value Boys=Girls	0.283	0.632	0.028	0.800	0.086	0.004	

Notes: Each cell represents the estimated impact and standard error from a different regression. The standard errors and *p*-values are clustered by school. The table shows standard errors in parenthesis and the *p*-values in brackets.

APPENDIX TABLES

Appendix Table 1: Characteristics of Inquiry and Problem Based Pedagogical Interventions

Country & year	Target Population	Grade	Didactic Materials	Teacher training	Teacher support	Reference
Math Interventions						
Argentina 2009	Public schools in disadvantaged communities in Tafí Viejo, Yerba Buena, and Cruz Alta in Tucumán, & southern Buenos Aires	4th grade	Workbook, calculator, rules, tables, games and figures	42 hours	Mentoring and training every other week	IDB, 2018a
Paraguay 2011	Preschools in Cordillera	Preschool	Workbook and audio lessons.	35 hours	Mentoring and training once a month	IDB, 2018c, IDB, 2018d
Paraguay 2013	Preschools in Cordillera	Preschool	Workbook and audio lessons	35 hours	Mentoring and training once a month	IDB, 2018e
Peru 2012	Preschools in Huancavelica, Angaraes and Ayacucho	Preschool	Math tools e.g., shapes, pictures, blocks, mirror, plastic tiles, and dice.	40 hours	Mentor visits once a month	IDB, 2018f
Belize 2015	Primary schools in Belize District	4th grade	Math tools such as tin frames, geometric solids, rods, etc.	29 hours	Mentor visits once a month	IDB, 2018b
Science Interventions						
Argentina 2009	Public schools in socioeconomically disadvantaged communities in Tafí Viejo, Yerba Buena, and Cruz Alta in Tucumán, and southern Buenos Aires	4th grade	Workbook and didactic materials	50 hours	Pedagogical and technical assistance	IDB, 2018a
Peru 2010	Public primary schools in Lima	3rd grade	LEGO kits	42 hours	Technical assistance and tutoring	IDB, 2018g
Peru 2012	Public primary schools in Lima	3rd grade	LEGO kits	73 hours	Technical assistance and tutoring	IDB, 2018g

Peru 2014	Students that perform in the 50% bottom in science scores in public primary schools in Lima.	3rd grade	Flipcharts	20 hours	None	IDB, 2018g
Belize 2015	Primary schools in Belize District	4th grade	Math tools such as tin frames, geometric solids, rods, etc.	29 hours	Mentor visits once a month	IDB, 2018b

Appendix Table 2. Experimental Design Characteristics

Country and year	School Sample Frame	Number of schools sampled	Schools allocated treatment	Stratifications for random assignment	Baseline collection dates	Follow up collection dates	Number Student s baseline	Number Students follow up
Argentina 2009	323	28	14	None	March, 2009	November 2009	1283	1126
Paraguay 2011	265	265	131	urban/rural, high/low school resources, & high/low school size	March 2011	Nov, Dec 2011	2907	2805
Paraguay 2013	265	262	129	urban/rural, high/low school resources, and high/low school size, and 1/2 sessions per day	March, April 2013	November 2013	3195	2888
Peru 2012	104	104	54	urban/rural, and geographic department	March, April, 2012	November 2012	2926	2400
Argentina 2009	323	42	28	None	March 2009	November 2009	2271	1927
Peru 2010	1203	106	53	urban/rural/metro, complete/multigrade, and and school size (small, medium or large).	April 2010	December 2010	2790	2392
Peru 2012	1203	104	52	urban/rural/metro, complete/multigrade, and and school size (small, medium or large).	March, April 2012	Nov, Dec 2012	2705	2401
Peru 2014	1217	48	Not applicable.	school and gender	May 2014	November 2014	1217	1127
Belize 2015	258	252	25	Urban/rural and funding (government or gov aided)	October, Nov 2014	May 2016	4713	4457

Appendix Table A3. Definition of Variables Used in the Analysis

Variable	Definition
Panel A. Individual Characteristics	
Math and Science Test Scores (std. deviations)	Designed to measure the ability of students to understand and apply key mathematical and scientific concepts adapted for each grade level and national curriculum. Standardized to mean zero and standard deviation of 1.
Student's age	Age of student in years.
Male	Equals 1 if student is male and 0 otherwise.
Mother completed secondary education	Equals 1 if mother has at least completed high school graduation and 0 otherwise
Both parents at home	Equals 1 if both parents living with student and 0 otherwise
Bilingual	Equals 1 if the child speaks Spanish and another language at home reported by parent and 0 otherwise
Asset Index (std. deviations)	Asset index created using principal component analysis to summarize information from the following variables: income per capita, number of people in the house, and housing floor, ceiling, and wall materials. Standardized to mean zero and standard deviation of 1.
Panel B. School and class characteristics	
Cohort size	Number of students enrolled in grade reported
Average class size	Cohort size divided by number classrooms
Teacher's education	Number of years of teacher education completed
Teacher experience	Number of years teacher has worked as a teacher.
Teacher is male	Equals 1 if the sex of the teacher is male and 0 otherwise.
Teacher's age in years	Age of teacher in years

Appendix Table A4: Baseline Descriptive Statistics and Tests of Balance between Treatment and Control groups

	Mathematics					Science				
	Argentina 2009	**Belize 2015**	**Paraguay 2011**	**Paraguay 2013**	**Peru 2012**	**Argentina 2009**	**Belize 2015**	**Peru 2010**	**Peru 2012**	**Peru 2014**
Test scores	-0.02	0	0	-0.04	0.04	-0.02	0	-0.03	-0.03	-0.03
	(-0.07)	(-0.14)	(-0.02)	(-0.06)	(-0.11)	(0.01)	(0.01)	(0.09)	(0.03)	(-0.02)
	[0.277]	[0.202]	[0.719]	[0.416]	[0.284]	[0.739]	[0.959]	[0.434]	[0.746]	[0.737]
Age	9.35	8.26	5	4.9	5	9.36	8.26	n.a.	8.02	8.19
	(-0.05)	(0.04)	(0.01)	(0.00)	(0.00)	(0)	(0.04)		(-0.05)	(0.02)
	[0.16]	[0.63]	[0.249]	[0.971]	[0.545]	[0.95]	[0.63]		[0.34]	[0.699]
Male	0.52	**0.5**	0.5	0.53	**0.57**	0.52	**0.5**	0.52	0.52	0.55
	(-0.02)	**(0.07)**	(-0.02)	(0.00)	**(-0.08)**	(-0.04)	**(0.07)**	(0)	(-0.02)	(0.00)
	[0.48]	**[0.001]**	[0.165]	[0.896]	**[0.003]**	[0.284]	**[0.001]**	[0.943]	[0.347]	
Bilingual	n.a.	n.a.	0.43	n.a.	0.12	0.14	n.a.	n.a.	0.06	0.89
			(-0.01)		(0.00)	(0.01)			(-0.01)	(-0.01)
			[0.678]		[0.984]	[0.827]			[0.56]	[0.462]
Asset Index	-0.04	-0.11	n.a.	n.a.	n.a.	-0.05	-0.11	n.a.	-0.01	n.a.
	(0.10)	(0.21)				(0.04)	(0.21)		(-0.04)	
	[0.080]	[0.1]				[0.76]	[0.110]		[0.573]	
Class size	**15.25**	23.42	15.36	17.13	21.45	**17.3**	23.42	23.36	22.48	23.81
	(-2.24)	(4.21)	(0.34)	(-0.13)	(2.02)	**(-2.05)**	(4.21)	(-1.52)	(-1.48)	(-0.06)
	[0.000]	[0.098]	[0.658]	[0.935]	[0.17]	**[0.083]**	[0.098]	[0.326]	[0.365]	[0.826]
Male teacher	n.a.	0.34	0.05	0.06	††	n.a.	0.34	n.a.	0.21	0.14
		(-0.05)	(0.04)	(0.02)			(-0.05)		(0.01)	(0)
		[0.712]	[0.205]	[0.586]			[0.712]		[0.932]	[0.748]
Teacher age	n.a.	35	35.25	37.14	n.a.	n.a.	35	n.a.	47.98	50.5
		(-2.59)	(0.75)	(0.04)			(-2.59)		(-0.61)	(-0.13)
		[0.278]	[0.314]	[0.953]			[0.278]		[0.725]	[0.741]

Notes: The table shows the control group mean, the average difference between the treatment and the control group means in parentheses, the *p*-values for a test differences in means with errors clustered at the school level, except for Argentina which are cluster bootstrapped. All estimates are based on baseline (pre-intervention) data. †† All teachers were female.

Appendix Table A5: Attrition Rates Between Baseline and Endline

Mathematics				Science				All
Argentina 2009	Paraguay 2011	Paraguay 2013	Peru 2012	Argentina 2009	Peru 2010	Peru 2012	Peru 2014	
Panel A: Attrition Rates								
0.13	0.03	0.08	0.21	0.17	0.16	0.11	0.08	0.08
(-0.01)	(0)	(0.02)	(-0.05)	(-0.04)***	(-0.03)	(0.01)	(-0.01)	(-0.01)
[0.493]	[0.660]	[0.100]	[0.348]	[0.003]	[0.107]	[0.419]	[0.609]	[0.217]

Notes: this table reports the means in the control group. Numbers in parenthesis show the average difference between the treatment and the control groups within strata. The numbers in brackets show the corresponding P-values for a test differences in means with errors clustered at the school level. The standard errors for Argentina are cluster bootstraped. *, **, *** indicate that the estimates coefficient is significantly statistically different from zero at the 0.10, 0.05 and 0.01 level, respectively.
Source: Authors' calculations.

Appendix Table A6: Baseline Descriptive Statistics and Tests of Balance between Treatment and Control Groups for Evaluation Sample

	Mathematics				Science			
	Argentina 2009	**Paraguay 2011**	**Paraguay 2013**	**Peru 2012**	**Argentina 2009**	**Peru 2010**	**Peru 2012**	**Peru 2014**
Test scores	0	0	0	0	0	0	0	0
	(-0.06)	(0.00)	(-0.08)	(-0.06)	(-0.01)	(0.08)	(0)	(-0.03)
	[0.338]	[0.942]	[0.241]	[0.541]	[0.825]	[0.473]	[0.979]	[0.514]
Age	9.35	5	4.91	5	9.36	n.a.	7.98	8.16
	(-0.03)	(0.01)	(0.00)	(0.00)	(0.02)		(-0.03)	(0.00)
	[0.439]	[0.21]	[0.828]	[0.434]	[0.808]		[0.583]	[0.977]
Male	0.52	0.5	0.53	**0.56**	0.52	0.51	0.52	0.56†
	(-0.02)	(-0.02)	(-0.01)	**(-0.07)**	(-0.04)	(0)	(-0.02)	(0.00)
	[0.52]	[0.159]	[0.702]	**[0.023]**	[0.336]	[0.898]	[0.372]	
Bilingual	n.a.	0.43	n.a.	0.13	0.14	n.a.	0.06	0.88
		(-0.01)		(0.00)	(0.01)		(-0.01)	(-0.01)
		[0.784]		[0.975]	[0.69]		[0.717]	[0.434]
Asset Index	-0.03	n.a.	n.a.	n.a.	-0.06	n.a.	0	n.a.
	(0.07)				(0.05)		(-0.05)	
	[0.252]				[0.687]		[0.55]	
Class size	**15.28**	15.35	17.13	21.34	**17.22**	23.64	22.6	23.81
	(-2.17)	(0.41)	(-0.29)	(1.6)	**(-1.97)**	(-1.71)	(-1.62)	(-0.01)
	[0.000]	[0.594]	[0.849]	[0.212]	**[0.093]**	[0.274]	[0.318]	[0.959]†
Male teacher	n.a.	0.05	0.06	††	n.a.	n.a.	0.2	0.13
		(0.04)	(0.01)				(0.01)	(0)
		[0.237]	[0.66]				[0.883]	[0.854]
Teacher age	n.a.	35.26	37.11	n.a.	n.a.	n.a.	48.01	50.42
		(0.71)	(-0.03)				(-0.56)	(-0.23)
		[0.343]	[0.973]				[0.753]	[0.61]

Notes: The table shows the control group mean, the average difference between the treatment and the control group means in parentheses, the *p*-values for a test differences in means with errors clustered at the school level. The standard errors for Argentina are cluster bootstrapped. All estimates are based on baseline (pre-intervention) data. †† All teachers were female.