Deficits in early childhood: Policy responses and their evaluation

Norbert Schady

October 2013

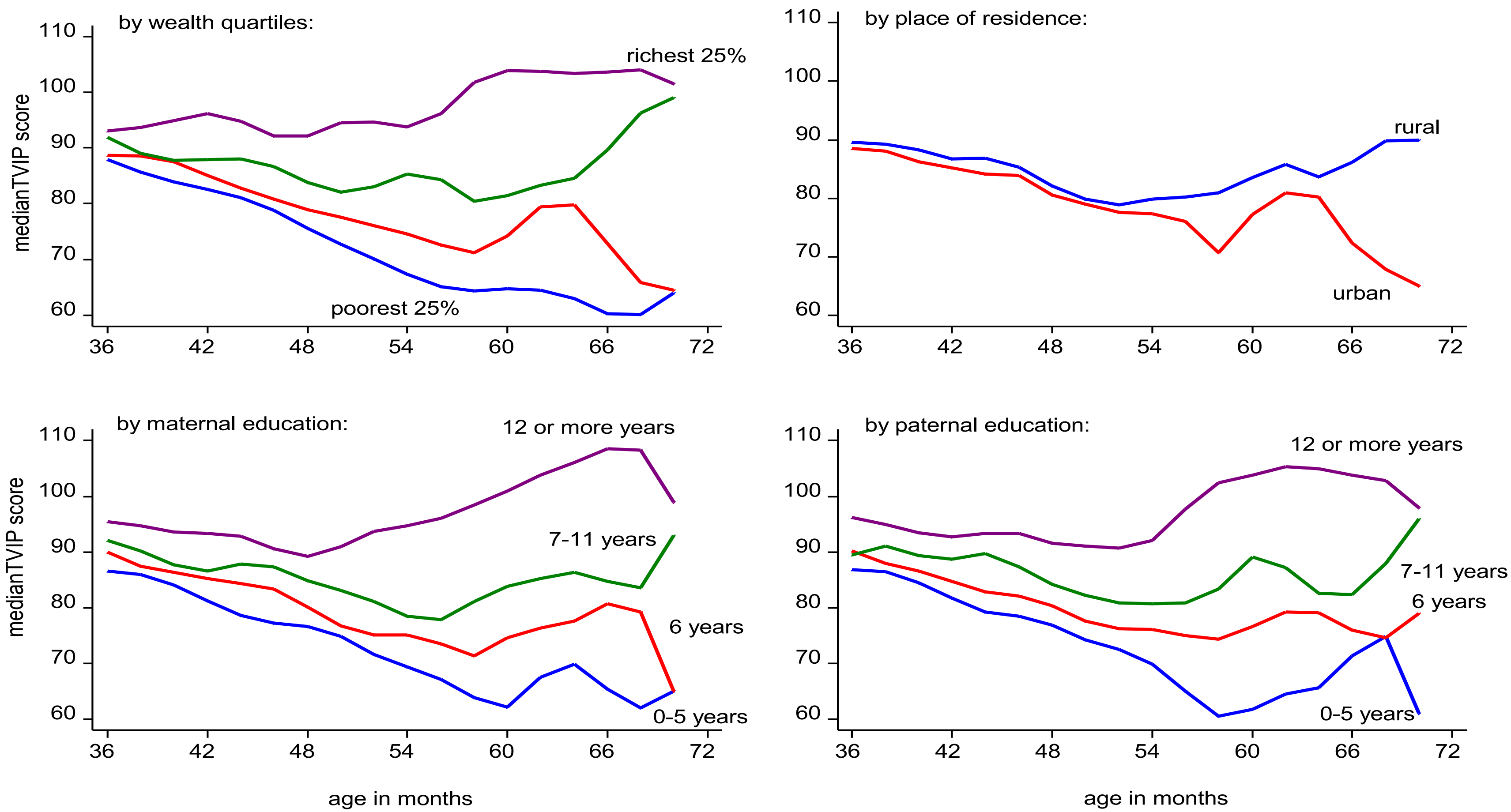
Motivation

- There is a growing consensus in the literature from medicine, child development, and economics that development in early childhood is critical in determining success in adulthood in a number of domains
- Long-term panels from developed and developing countries that have followed children from early ages into adulthood show that children with poor levels of nutrition, inadequate cognitive development, and low levels of socio-emotional development tend to do badly in school, have higher levels of unemployment, earn lower wages (even controlling for years of schooling), have a higher incidence of teenage pregnancy, are more likely to use drugs, and are more likely to be involved in criminal activities
  - Reviews include Almond and Currie (2011) for the US, Engle et al. (2007; 2011) and Grantham-McGregor et al. (2007) for developing countries

Motivation

- In Latin America and the Caribbean, there is also a great deal of policy experimentation and (some) credible impact evaluation of programs that attempt to improve outcomes in early childhood
- Before children enter the formal education system ("prevention")
  - Center-based care for young children
  - Parenting programs
  - Cash transfer programs
- Once children are age-eligible for the formal education system ("early remediation")
  - Expanding access
  - Improving quality

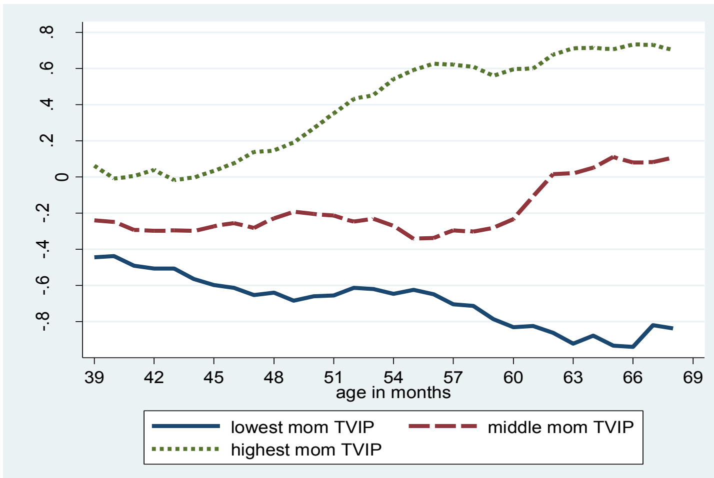
SES gradients in cognitive development



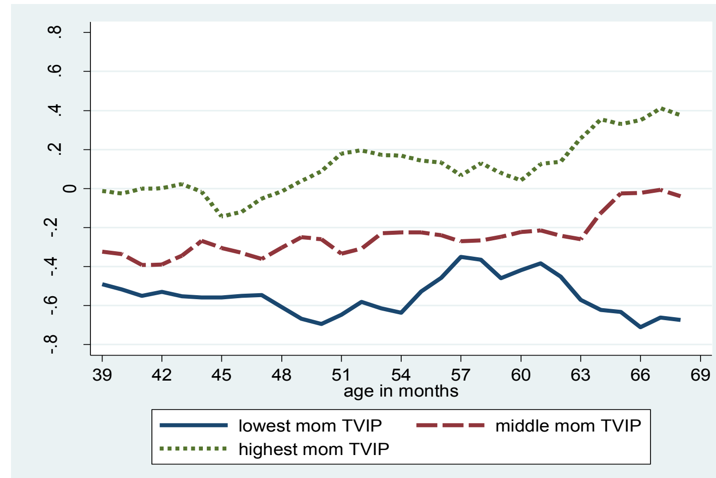
Source: Paxson and Schady (JHR, 2007)

SES gradients in cognitive development

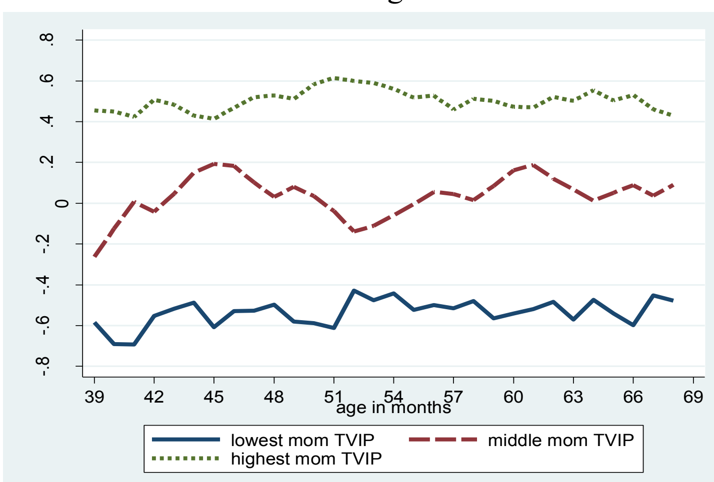
Vocabulary



Memory

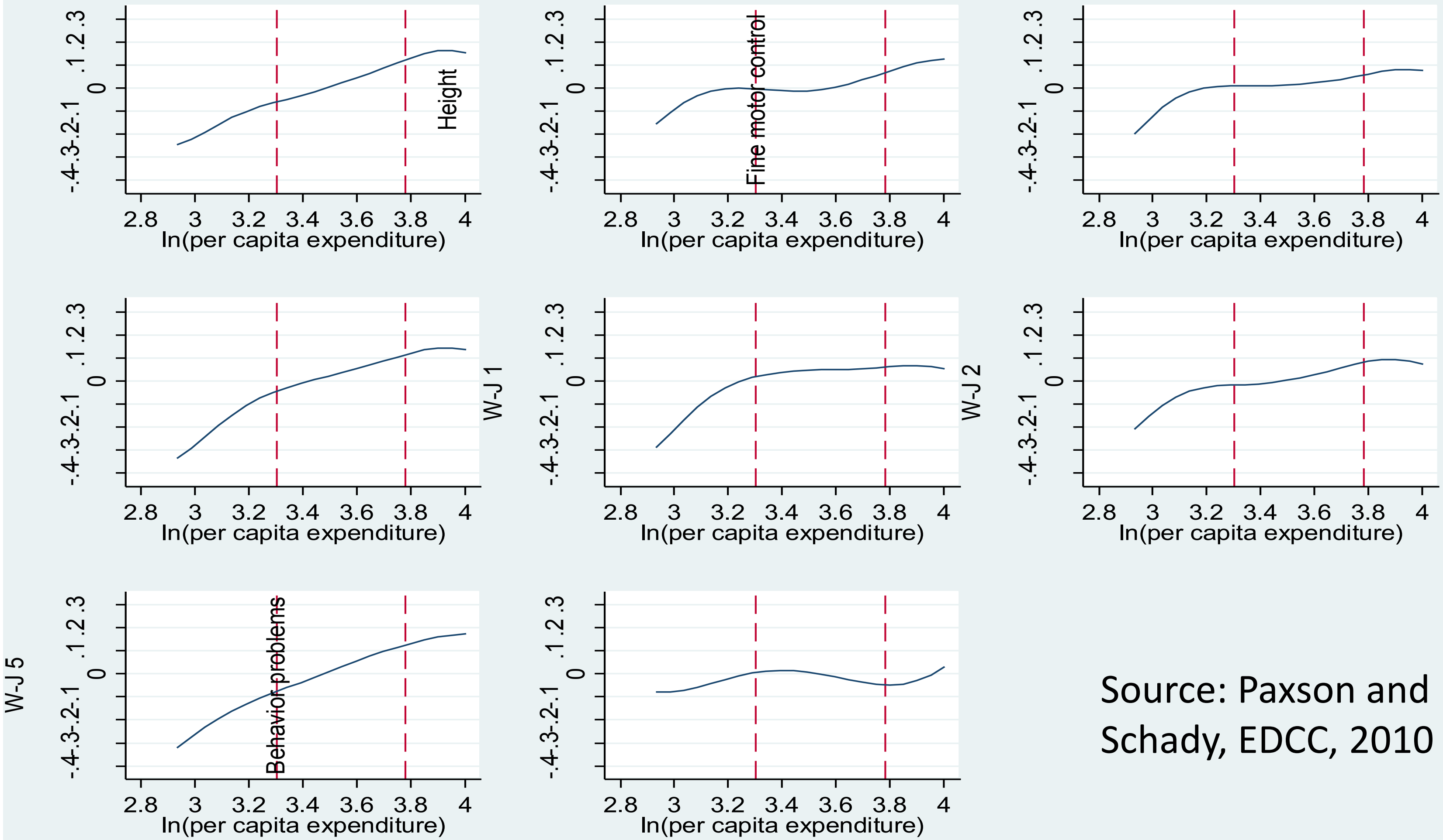


Visual integration



Source: Schady (2011)

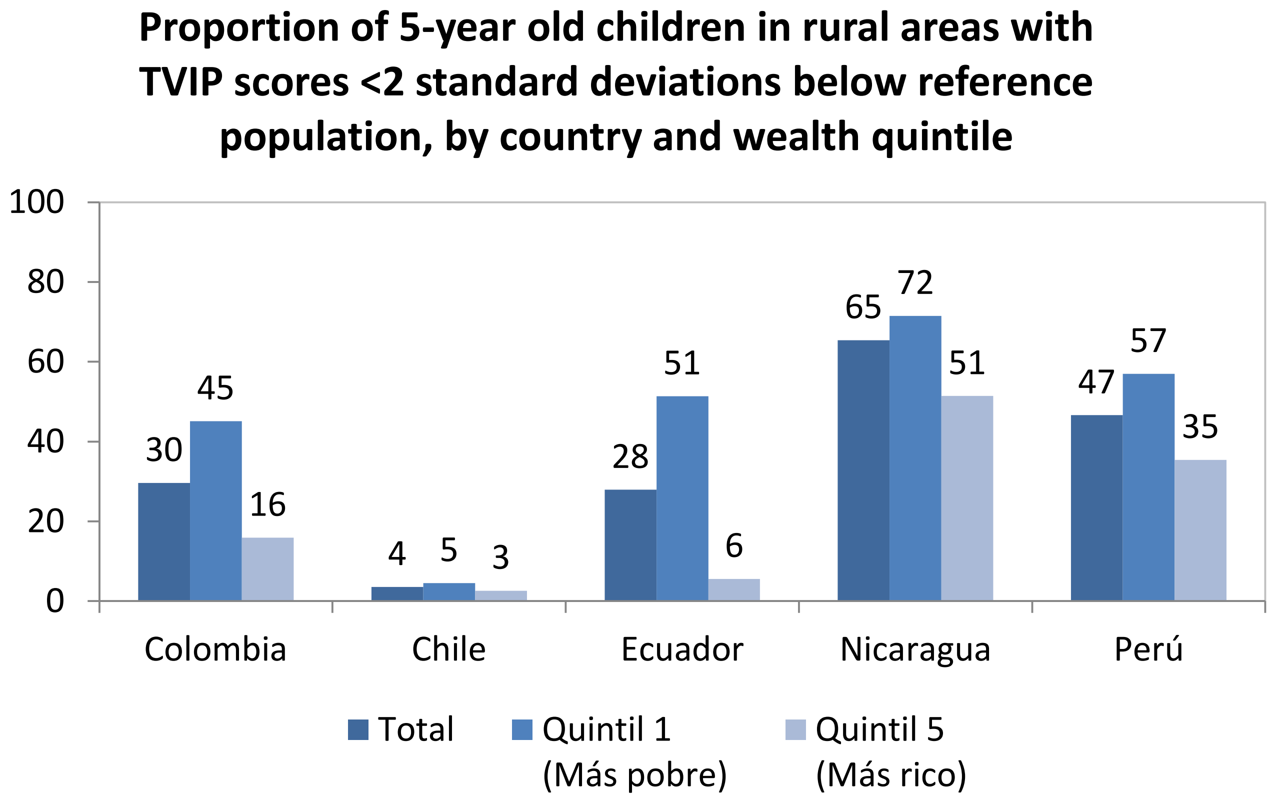
SES gradients in cognitive development



Source: Paxson and Schady, EDCC, 2010

- All outcomes are converted to z-scores
- Reversed sign if necessary so that higher values correspond to better outcomes

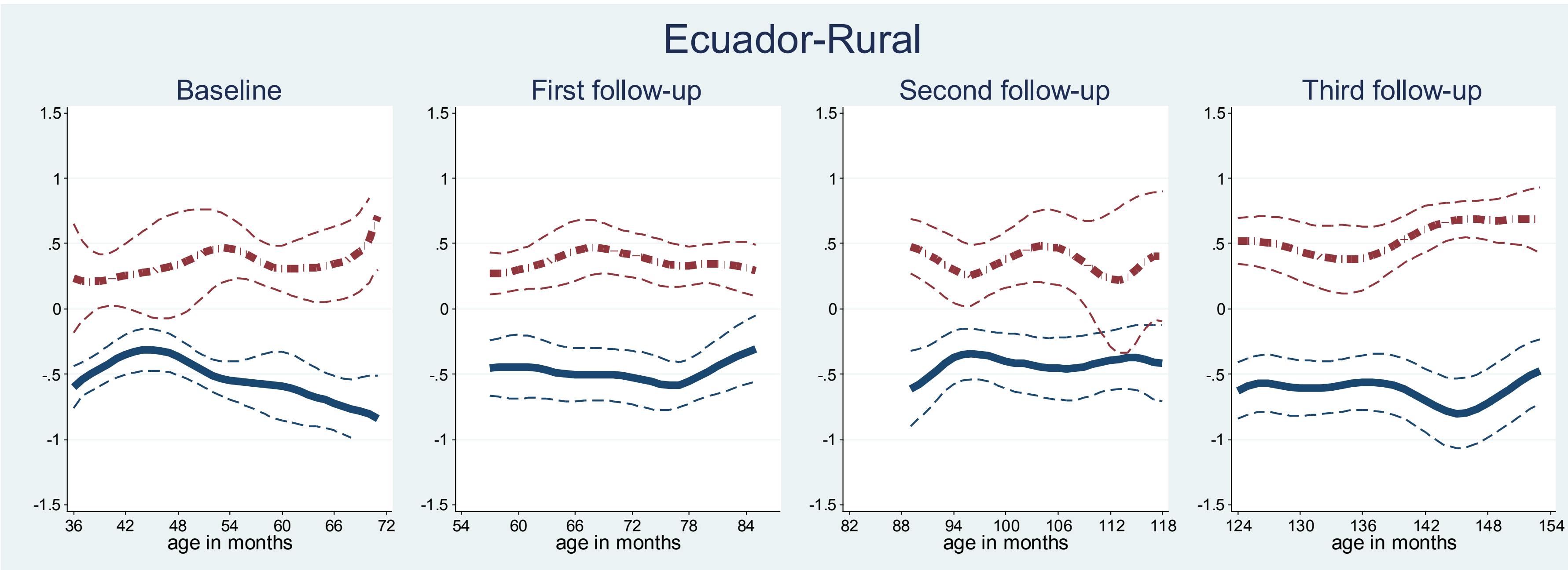
SES gradients in cognitive development



Source: Own calculations, based on data in Schady et al. (2012)

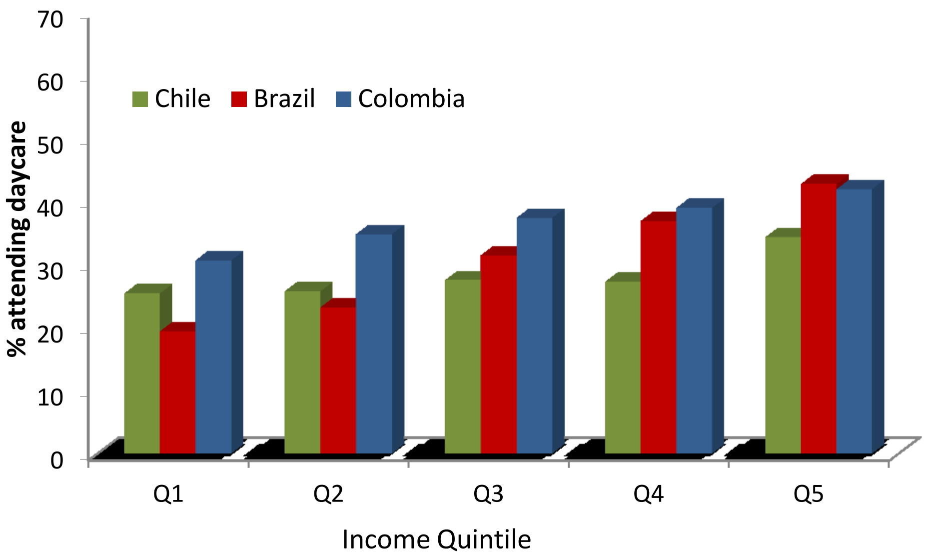
SES gradients in cognitive development

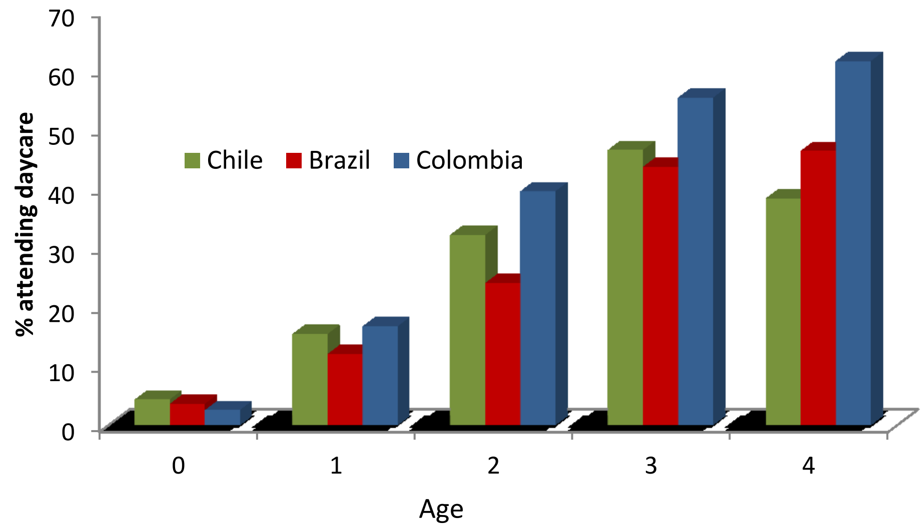
.



Source: Schady et al. (2012)

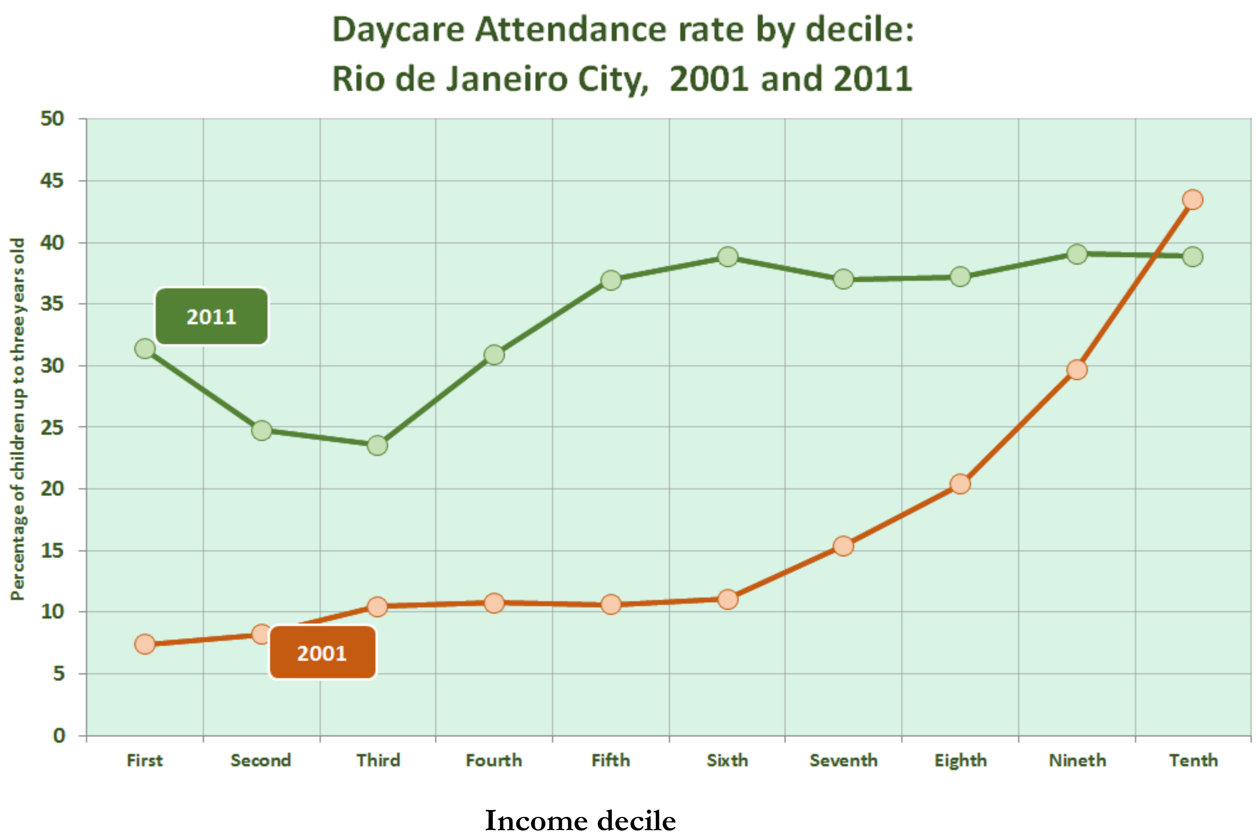
Center-based care, Brazil, Chile, Colombia





Source: Own calculations based on data from Brazil (PNAD 2012), Chile (CASEN 2011) and Colombia ENDS (2010)

Center-based care, Rio de Janeiro



In **Rio**, the expansion in coverage of crèches has been dramatic, especially among poor households

Source: Paes de Barros (2013)

Center-based care, Rio de Janeiro

<u>Experimental impact evaluation</u> (Barros et al. 2013)

- Excess demand for crèches (2.1 applications for each spot)
- Random assignment (crèche by crèche) of spots (2007)
  - 243 separate experiments
  - Verification of random assignment using administrative data
  - Household survey in 2008
    - Increase of ~17 percentage points in female labor supply (participation) and increases of 2-0-25% in household income
  - Household survey, including tests for children, carried out in late 2012: includes three tests of cognitive development, three tests of executive function, mother-reported measures of child behavior and development
    - Increase in employment and income remains
    - Improvements of ~0.2 a 0.3 standard deviations in nutritional status, cognitive development, executive function (although estimates somewhat noisy)
  - Note: program is expensive (~US $250 per child per month) but the increase in income is also substantial in magnitude (US $230 per treated family)

Parenting and early stimulation

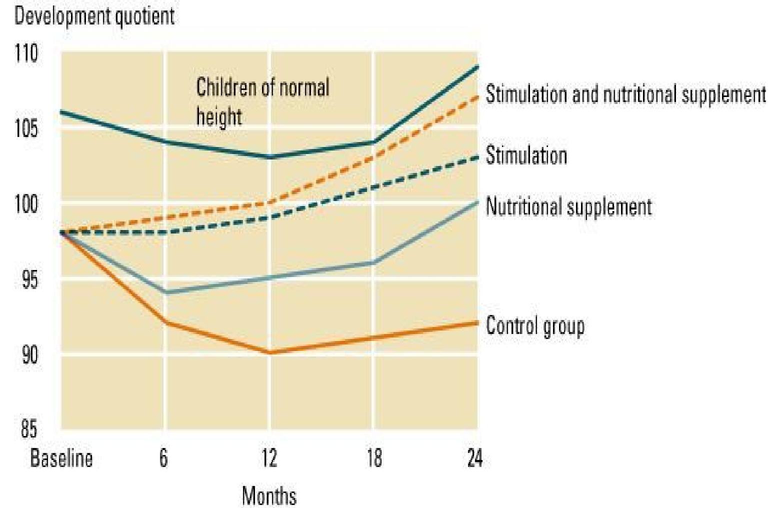
The Jamaican study

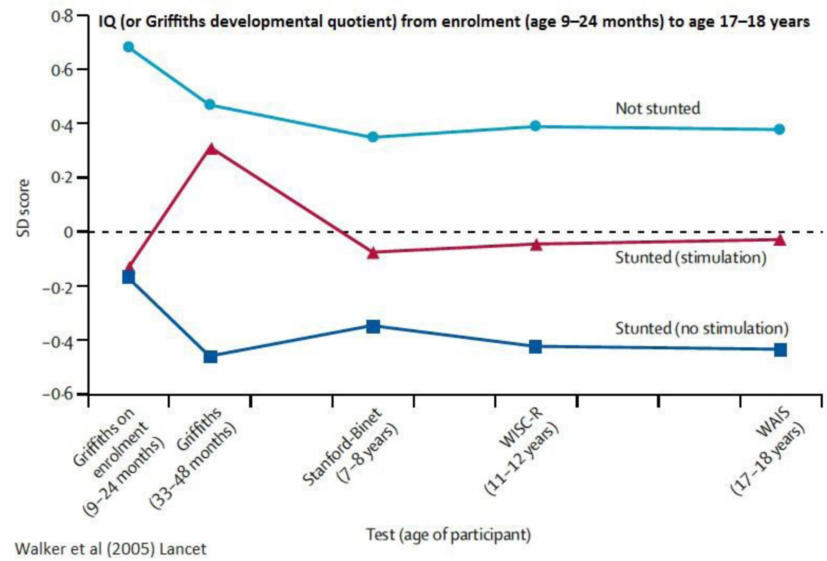
Random assignment to four treatment arms + control group:

- Group 1: Nutritional supplement
- Group 2: Stimulation
- Group 3: Nutritional supplement + stimulation

Very small sample sizes (n=129) (efficacy trial) but very long follow-up

In the most recent follow-ups (Walker et al. 2011; Gertler et al. 2012), children who received stimulation had higher IQ (about 0.6 standard deviations), better performance on tests of mathematics and reading, higher levels of completed schooling (about one-third more years), lower levels of depression, lower levels of involvement in violent criminal activity and better labor market outcomes





Source: Walker, Grantham-McGregor and co-authors, various years

Parenting and early stimulation

- Jamaican curriculum, adapted to local circumstances
- "Madres líderes" elected by community, training and supervision by professionals
- Weekly home visits

Impact of home visits in Colombia Colombia							
		12-18 months at outset			18-24 months at outset		
		S	N	S+N	S	N	S+N
Cognitive development		0.16* (0.11)	-0.02 (0.09)	0.22** (0.11)	0.37** (0.11)	0.13** (0.09)	0.22** (0.10)
Receptive language		0.10 (0.10)	-0.14 (0.11)	0.10 (0.11)	0.39** (0.13)	0.24* (0.12)	0.22* (0.12)
Expressive language		-0.04 (0.11)	-0.07 (0.13)	-0.08 (0.12)	0.28** (0.11)	0.23** (0.11)	0.21** (0.11)
Fuente: Attanasio et al. (2013). S=Stimulation; N=micronutrient supplement							

Cash transfers in Ecuador and Nicaragua

Ecuador Bono de Desarrollo Humano (Paxson and Schady 2010)

- Cash transfer to women, equivalent to ~ 10 percent of consumption for mean recipient household
- Conditionality announced, but never monitored
- Information campaign, stressing that parents who received BDH transfers were responsible for the education and health of their children

Nicaragua *Atención a Crisis* (Macours et al. 2012)

- Cash transfer to women, equivalent to ~ 15 percent of consumption for mean recipient household
- Conditionality announced, monitored in education, but not in health
- Widespread, intensive social marketing campaign: transfers were intended to improve the diversity and nutrient content of children's diets and to buy school material

- Both programs probably best thought of as unconditional cash transfer, with cash given to women, and social marketing to influence use of resources

Cash transfers in Ecuador and Nicaragua

Ecuador

- No baseline differences between treatment and control groups
- High level of compliance with experimental design
  - 75 percent of treatment group received transfers
  - 3.7 percent of control group received transfers
- Households received transfer for 17 months, on average
- Low levels of attrition: 5.9 percent
  - Uncorrelated with treatment status
- Poor households:
  - 34 percent < 1 US $ per capita per day
  - 93 percent < 2 US $ per capita per day
- Did not collect data on log pce
- Paper analyzes two waves of data

Nicaragua

- No baseline differences between treatment and control groups
- High level of compliance with experimental design
  - 100 percent of treatment group received transfers
  - < 1 percent of control group received transfers
- Households received transfer for 9 months, on average
- Very low levels of attrition: 1.3 percent
  - Uncorrelated with treatment status
- Extremely poor households
  - 82 percent < 1 US $ per capita per day
- Collected data on log pce
- Paper analyzes three waves of data
  - **Transfers discontinued after second wave**

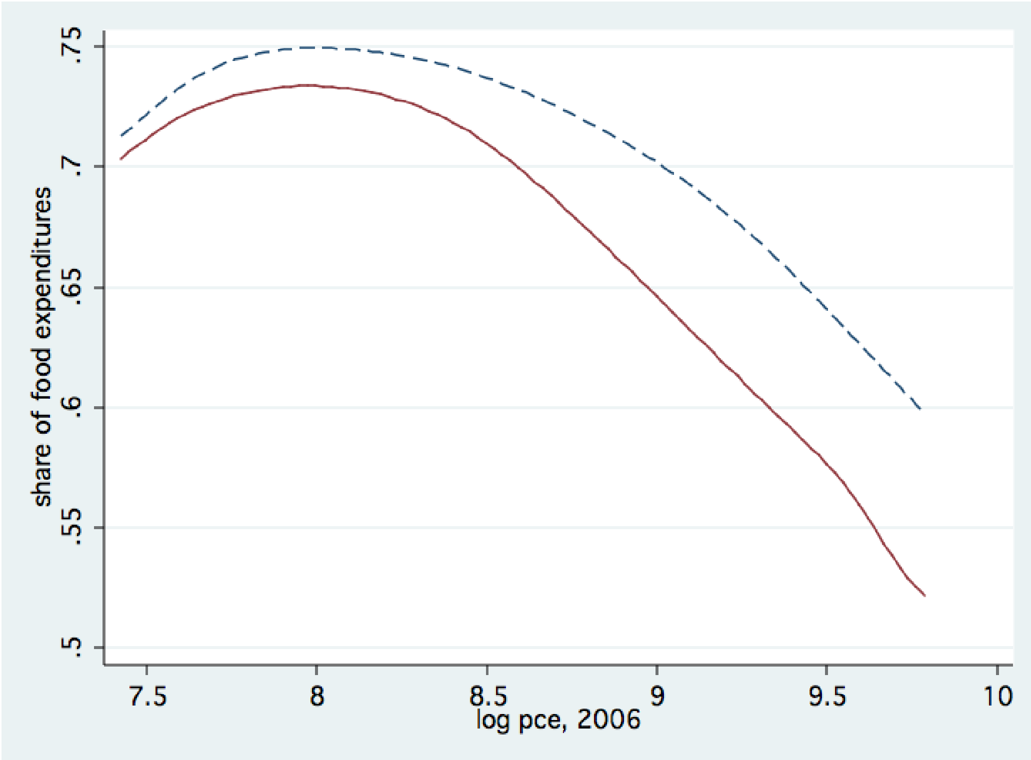
Cash transfers: Ecuador

	Full sample		Bottom quartile
Hemoglobin	0.074 (0.112)		0.293*** (0.113)
Height-for-age	-0.004 (0.056)		0.022 (0.084)
Fine motor control	0.100 (0.068)		0.159 (0.101)
Receptive vocabulary (TVIP)	0.005 (0.098)		0.137 (0.129)
Long-term memory (Martians)	0.141 (0.092)		0.173* (0.096)
Short-term memory (Digit span)	-0.019 (0.100)		0.079 (0.143)
Visual integration	0.053 (0.095)		0.256 (0.160)
Behavior problems	0.066 (0.091)		0.240 (0.147)
Av. effects: Physical measures	0.057 (0.047)		0.158 (0.060)***
Av. effects: Cognitive & Behavioral	0.049 (0.066)		0.177 (0.094)**
Av. Effects: All measures	0.052 (0.052)		0.170 (0.074)**
Note: Coefficients and standard errors. Children age 36-83 months at follow-up. Mean effect sizes estimated by Seemingly Unrelated Regression (SUR)			

Cash transfers: Ecuador

	2006		2008
Denver: Language	0.14*** (0.05)		0.09** (0.05)
Denver: Social personal	0.13*** (0.05)		0.10 (0.05)
Denver: Gross motor	-0.01 (0.05)		0.10 (0.06)
Denver: Fine motor	0.04 (0.06)		0.16*** (0.04)
Height	0.07** (0.03)		0.05 (0.03)
Weight	0.04 (0.04)		0.03 (0.04)
Leg motor	0.13* (0.08)		0.14 (0.05)
Behavior	-0.05 (0.08)		0.02 (0.06)
Receptive vocabulary (TVIP)	0.23*** (0.06)		0.09 (0.08)
Long-term memory (Martians)			0.09** (0.04)
Short-term memory (Digit span)	0.16*** (0.04)		0.11** (0.05)
Av. effects: Physical measures	0.05 (0.04)		0.07 (0.03)**
Av. effects: Cognitive & Behavioral	0.12*** (0.03)		0.08** (0.03)
Av. Effects: All measures	0.09** (0.03)		0.08*** (0.03)
Note: Coefficients and p-values; sample limited to children tested in both survey waves			

Is it the cash or something else? Food Engel curves



- Nicaragua: Three different treatment arms + control
- One group received substantially larger transfer (26% of mean pce, as opposed to 15% of mean pce), and had 17% higher consumption
- This group did not have improved child development outcomes
- Magnitude of effects also appears "too large" when compared with the cross-sectional elasticity of child development with respect to income or consumption l

Access to kindergarten

Preschool has been shown to have benefits in *some* countries in Latin America:

- In Uruguay, attending preschool leads to ~0.8 additional years of completed schooling by age 15 (Berlinski et al. 2008)
  - Benefit-cost ratios of preschool of 3.2 using a discount rate of 10%
  - Conditional on attending preschool at age 5, preschool attendance at ages 3 or 4 has no discernible benefits—at least, in terms of the years of schooling completed
- In Argentina, one year of preschool increases test scores in language and mathematics in third grade by ~0.23 sds (Berlinski et al. 2009)
  - Preschool participants also have fewer behavior problems, are more likely to pay attention in class, and are more likely to participate, as reported by their third-grade teachers
- In Guatemala, a preschool construction program had no effect on primary school attendance, and modest effects (1.5% points, from a mean of 22%) on the proportion of children with age-appropriate grade progress (Bastos et al. 2012)

Can better teachers compensate for early deficits? Evidence from a randomized experiment in Ecuador

Pedro Carneiro (UCL)

Yyannu Cruz-Aguayo (IDB)

Norbert Schady (IDB)

October 2013

Ecuador study design

- Take random sample of 204 schools with 2 or more kindergarten classes in the coastal region of the country
- Randomly assign an entering cohort of ~ 15,000 children to classrooms at start of 2012/2013 school year (May-January)
  - Note that these are in effect 204 separate experiments

Outline

- Some basic facts about Ecuador
- Identification challenges and study design
- Descriptive results: The CLASS
- Classroom effects
- Teacher characteristics, behaviors, and student learning outcomes

Ecuador: Some basic facts

Table 1: Preschool attendance rates, children aged 5 years

	Late 1990					Late 2000				
Country	**Year**	**Mean**	**Q1**	**Q5**	**Gap Q5-Q1**	**Year**	**Mean**	**Q1**	**Q5**	**Gap Q5-Q1**
Argentina	1999	75.3	65.0	91.9	26.9	2010	95.5	93.3	99.0	5.7
Brazil	1999	60.2	50.9	89.2	38.2	2009	84.4	79.8	95.7	15.9
Chile	1998	66.6	56.1	77.4	21.3	2009	90.3	87.2	95.9	8.7
Colombia	1999	67.3	51.9	86.4	34.5	2010	78.9	67.2	95.5	28.3
Costa Rica	1999	34.3	27.3	49.3	22.0	2010	71.8	66.2	88.0	21.8
Dom. Rep.	2000	66.4	52.7	82.9	30.2	2010	83.8	79.6	91.0	11.3
Ecuador	**2000**	**70.2**	**62.7**	**78.7**	**16.1**	**2012**	**93.2**	**89.7**	**100.0**	**10.3**
Honduras	1999	34.4	27.2	56.6	29.4	2010	69.8	66.3	88.4	22.1
Mexico	1998	76.6	59.4	95.1	35.7	2010	96.4	92.4	100.0	7.6
Nicaragua	1998	56.2	38.0	62.3	24.3	2010	72.6	66.0	94.4	28.4
Panama	1999	64.2	46.9	89.7	42.8	2010	80.7	68.4	93.8	25.4
Peru	1999	79.3	67.1	92.9	25.8	2010	91.6	82.7	99.5	16.8
Paraguay	1999	43.9	34.5	67.2	32.7	2010	68.5	48.3	79.4	31.2
El Salvador	1999	44.9	26.1	82.7	56.6	2010	66.5	56.6	90.8	34.2
Uruguay	1999	90.5	83.7	99.2	15.5	2010	97.3	96.5	98.2	1.7
Venezuela	1999	83.1	73.4	91.8	18.5	2010	91.4	89.2	96.7	7.6

Source: Own calculations, based on data from households surveys

Ecuador: Some basic facts

- Population: 15.2 million
- GDP per capita: US $ 8,393 (in PPP dollars)
- Average annual growth in GDP per capita 2003-2012: 3.1%
- Gini coefficient: 0.47
  - US: 0.41
  - Sweden: 0.25
  - Average OECD (excluding Mexico and Chile): 0.32
- Mean years of completed schooling of population 25+
  - Men: 8.7
  - Women: 8.4
- Life expectancy: 75.8 years
- Fraction of children stunted (low height for age): 18.1%

Ecuador: Some basic facts

Table 2: Performance in math, SERCE (2006) third grade				
Country	**Low**	**Medium**	**High**	**Advanced**
Argentina	43.23	31.13	15.17	10.47
Brazil	46.87	26.74	14.32	12.07
Chile	33.00	33.60	19.37	14.02
Colombia	47.17	33.19	12.97	6.67
Costa Rica	27.06	37.00	22.30	13.65
Ecuador	**59.82**	**28.12**	**7.91**	**4.14**
El Salvador	55.31	31.80	9.25	3.64
Guatemala	67.40	25.07	5.46	2.08
Mexico	34.00	30.70	19.71	15.59
Nicaragua	60.05	30.50	7.49	1.97
Panama	65.67	25.15	6.42	2.75
Paraguay	53.75	25.50	11.56	9.20
Perú	60.66	25.95	8.61	4.77
Dominican Republic	90.55	8.49	0.84	0.13
Uruguay	31.73	30.03	19.29	23.11

Source: LLERCE (2008)

Ecuador: Some basic facts

- 3.56 million children age 5-18 in education system (80 percent in public, 20 percent in private)
- 155,819 public sector teachers
  - 109,877 with tenure, 45,942 on a contract basis
  - 18,277 public schools, 5,078 private
- Teachers selected for tenure on basis of "*Concurso de Méritos y Oposiciones*" test
  - In earlier work (Cruz-Aguayo, Ibarraran, Schady 2013) we have shown that performance on the test is not predictive of child learning
- 32 universities offer teaching majors
- Teacher salary scale based to an overwhelming extent on seniority: No pay for performance
- In-service training done by institution within Ministry of Education

Outline

- Some basic facts about Ecuador
- Identification challenges and study design
- Descriptive results: The CLASS
- Classroom effects
- Teacher characteristics, behaviors, and student learning outcomes

Identification challenges

1. <u>Challenge 1</u>: In general, children are not assigned to teachers randomly, and the endogeneity of the match between teachers and students raises difficult estimation issues
   – If worse children are assigned to better teachers—as might be the case, for example, if headmasters are trying to maximize learning among weakest children, or trying to equalize outcomes—then estimates of "teacher effects" will generally be biased down
   – Conversely, if better children are generally assigned to better teachers—as might be the case, for example, if better teachers have more lobbying capacity, or if parents of more capable children are more successful placing their children with better teachers—then estimates of "teacher effects" will be biased up

Identification challenges

- In the economics literature, researchers have attempted to deal with this identification challenge estimating Value Added Models (VAM) or with a variety of fixed effect models
  - But there is controversy about the extent to which these models address the identification challenge (Krueger 1999; Andrabi et al. 2011; Rothstein 2010; Chetty et al. 2013; Kane and Staiger 2008)
- An alternative which convincingly deals with this challenge is to randomly assign children to teachers

Evidence from random assignment

- Best known and very influential study is **Project STAR** in Tennessee (Krueger 1999; Krueger and Whitmore 2001; Chetty et al. 2011)
  - Children randomly assigned to small classes (13-17 children), large classes (22-25 children), large classes with teacher aide
    - Children were supposed to remain in assigned classes for 4 years (k through 3rd grade)
  - Some limitations:
    - No baseline data on child development
    - Problems of non-compliance with random assignment, especially after kindergarten
    - Very high levels of attrition of students: ~50% over three years

Evidence from random assignment

- The "Measuring Effective Teaching" (MET) study currently underway (Kane and Staiger 2012)
    - Only covers teachers who *volunteered* into the study
    - A great deal of contamination of random assignment by both teachers and children: ~approximately 50% of students violated random assignment
        - Substantial variation across sites, compliance rate of 66% in Dallas, 27% in Memphis

Identification challenges

2. Challenge 2: The observable characteristics of teachers—years of experience, education, how teachers are hired—generally explain very little of the observed variation in teacher effects

- "Consistent with much of the previous literature, the STAR data suggest that measured teacher characteristics explain relatively little of student achievement on standardized tests." (Krueger 1999)
- "The results demonstrate quite clearly that the observable school and teacher characteristics explain little of the between-classroom variation in achievement growth despite the fact that a substantial share of the overall achievement growth occurs between teachers." (Rivkin et al. 2005)

- Recent research has attempted to collect data on a variety of teacher practices, in particular using different classroom observation tools (including Danielson 1996; Pianta and Hamre 2009; Pianta 2011)

Ecuador study design

- Random assignment
- Collect data on children at the beginning of the school year (TVIP, age, gender)
- Collect a short household survey (parental education, "wealth", home environment, stimulation, perceptions of school and teacher quality)
- Collect very rich data on teachers
  - "Traditional" measures of observable characteristics (experience, education, gender, whether tenured or hired on a contract basis, whether has received in-service training)
  - The CLASS (Pianta 2011; Pianta and Hamre 2009)
    - All 451 teachers filmed for an entire school day during 2012/2013 school year
    - Most teachers also filmed during 2011/2012 school year
- Collect very rich data on students at end of 2012/2013 school year: 4 tests of language, 4 or math, 4 of executive function

Baseline characteristics

Table 3. Descriptive statistics: children and families						
	Cerrando Brechas Sample			National Sample		
Age of child (months)	66.97	(5.2)	[14883]	66.31	(5.99)	[1576]
Median		67			72	
10th percentile		62			60	
90th percentile		72			72	
Proportion Female	0.49	(0.5)	[14996]	0.48	(0.5)	[1576]
TVIP	82.86	(15.88)	[13621]			
Median		81				
10th percentile		63				
90th percentile		105				
Age of parents						
Mother's Age	30.23	(6.57)	[13564]	34.16	(7.13)	[1498]
Median		29			34	
10th percentile		23			25	
90th percentile		39			43	
Father's Age	34.55	(7.89)	[10569]	38.23	(8.97)	[1327]
Median		33			37	
10th percentile		26			27	
90th percentile		45			50	

Source: Own calculations from Cerrando Brechas and ENEMDU (2012).

Notes: National sample restricted to 5-6 year-old children of the head of household. Mean, standard deviation (in parentheses) and number of observations [in square brackets]

Baseline characteristics

Table 4. Descriptive statistics: children and families						
	Cerrando Brechas Sample			National Sample		
Education of parents						
Mother's Education	8.78	(3.81)	[13554]	8.77	(4.46)	[1498]
Median		9			9	
10th percentile		5			3	
90th percentile		12			15	
Father's Education	8.49	(3.84)	[10544]	9.08	(4.47)	[1327]
Median		8			9	
10th percentile		5			4	
90th percentile		12			16	
Housing characteristics						
Proportion with access to piped water	0.83	(0.38)	[14330]	0.81	(0.39)	[1504]
Proportion with access to sewage	0.46	(0.5)	[14330]	0.61	(0.49)	[1517]

Source: Own calculations from Cerrando Brechas and ENEMDU (2012).
Notes: National sample restricted to 5-6 year-old children of the head of household. Mean, standard deviation (in parentheses) and number of observations [in square brackets]

Battery of tests

- 12 tests (TVIP, Woodcock-Johnson battery, Stroop test, multi-dimensional card sort)
- Piloted and validated in Ecuador
- Approximately 40 minutes per child

3 areas:

- **<u>Language and early literacy</u>**
  1. Letter & word id
  2. Id of first sound of words
  3. Receptive vocabulary (TVIP)
  4. Oral comprehension
- **<u>Math</u>**
  1. Number ID
  2. Block rotation
  3. Numeric series
  4. Applied problems
- **<u>Executive function</u>**
  1. Inhibitory control
  2. Attention control
  3. Cognitive flexibility
  4. Working memory

Very low levels of performance

.

Letter word ID

A

la **pan** **de** **tren**

C

O

t

At the end of kindergarten:

- On average, children can identify 2 letters
- Almost half the children in the sample cannot recognize a single letter or word

Very low levels of performance

Number ID

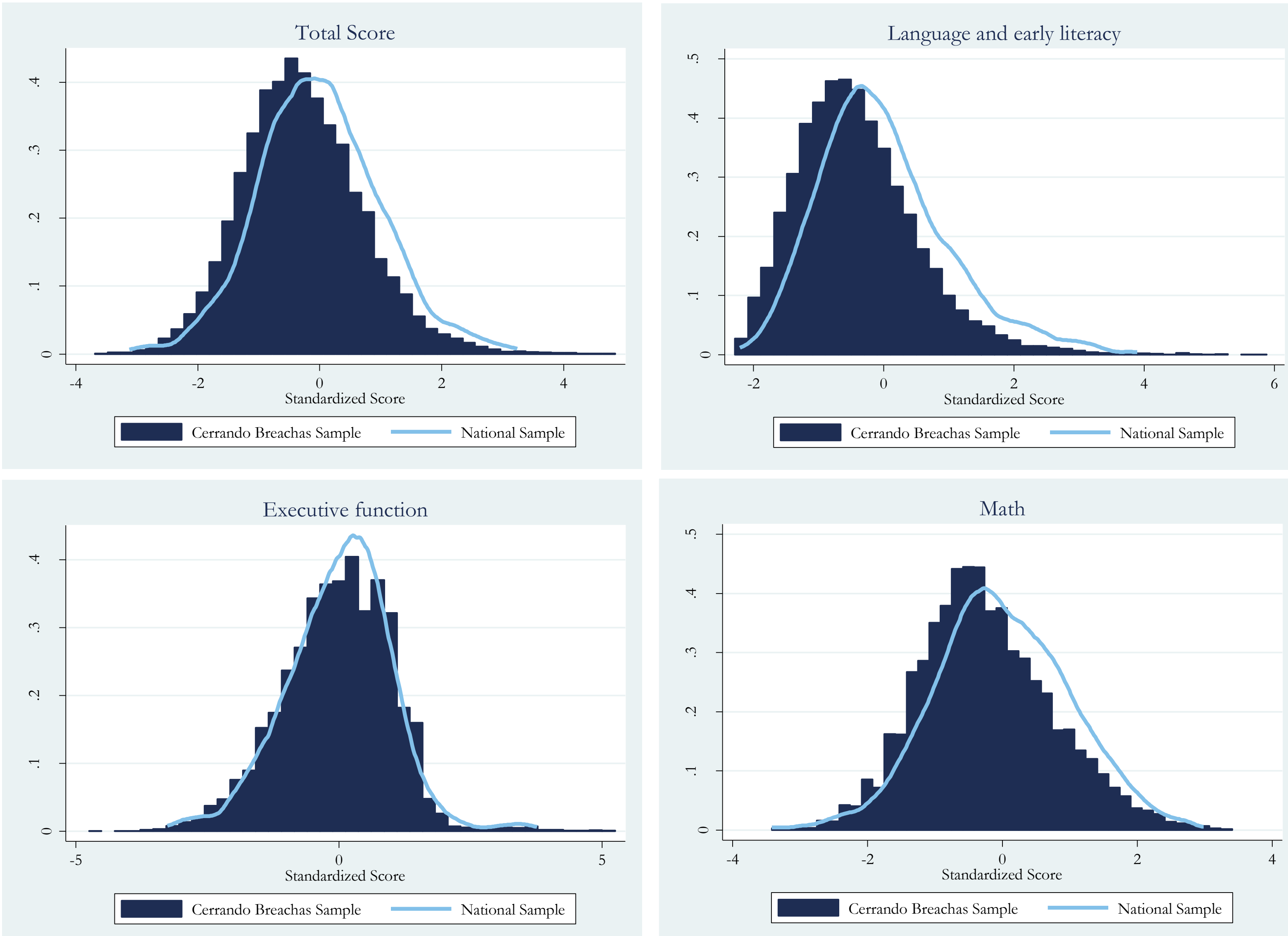
.

4
2
7
12

At the end of kindergarten:

- On average, children recognize 4 numbers
- More than 10 percent of children do not recognize a single number

Very low levels of performance



Baseline characteristics

Table 5: Descriptive statistics: Teachers						
	Cerrando Brechas			National Sample		
Teacher Age	42.23	(9.58)	[448]	43.49	(10.19)	[202]
Proportion Female	0.99	(0.10)	[450]	0.98	(0.14)	[202]
Years of experience	14.91	(8.88)	[450]	17.78	(10.56)	[202]
Years teaching at current school	7.43	(7.54)	[450]	10.3	(8.81)	[202]
Tenured	0.64	(0.48)	[450]	0.86	(0.35)	[202]
Proportion who have had in-service training (last 3 years)	0.69	(0.46)	[450]	0.85	(0.36)	[202]
Class size (2012/13 school year)	33.23	(7.83)	[451]	31.37	(8.67)	[200]

Note: Mean, standard deviation (in parentheses) and number of observations [in square brackets]

What is the CLASS?

CLASS: Classroom Assessment Scoring Scheme

•Developed by Robert Pianta and others at the University of Virginia (Pianta 2011; Pianta and Hamre 2009)

•Tool of classroom observation widely used and validated.

•Measures the quality of teacher's interaction with her students in three domains, on a scale from 1 to 7.

- Emotional Support: encompasses positive or negative climate in the classroom, teacher sensitivity, and regard for student perspectives.
- Classroom Organization: encompasses behavior management, productivity, instructional learning formats.
- Instructional Support: encompasses concept development, quality of feedback, language modeling.

Behavior Management

Encompasses the teacher's ability to provide clear behavioral expectations and use effective methods to prevent and redirect misbehavior.

	Low (1,2)	Mid (3,4,5)	High (6,7)
Clear Behavior Expectations ▪ Clear expectations ▪ Consistency ▪ Clarity of rules	Rules and expectations are absent, unclear, or inconsistently enforced.	Rules and expectations may be stated clearly, but are inconsistently enforced.	Rules and expectations for behavior are clear and are consistently enforced.
Proactive ▪ Anticipates problem behavior or escalation ▪ Rarely reactive ▪ Monitoring	Teacher is reactive and monitoring is absent or ineffective.	Teacher uses a mix of proactive and reactive responses; sometimes monitors but at other times misses early indicators of problems.	Teacher is consistently proactive and monitors effectively to prevent problems from developing.
Redirection of Misbehavior ▪ Effectively reduces misbehavior ▪ Attention to the positive ▪ Uses subtle cues to redirect ▪ Efficient	Attempts to redirect misbehavior are ineffective; teacher rarely focuses on positives or uses subtle cues. As a result, misbehavior continues/escalates and takes time away from learning.	Some attempts to redirect misbehavior are effective; teacher sometimes focuses on positives and uses subtle cues. As a result, there are few times when misbehavior continue/escalate or takes time away from learning.	Teacher effectively redirects misbehavior by focusing on positives and making use of subtle cues. Behavior management does not take time away from learning.
Student Behavior ▪ Frequent compliance ▪ Little aggression & defiance	There are frequent instances of misbehavior in the classroom.	There are periodic episodes of misbehavior in the classroom.	There are few, if any, instances of student misbehavior in the classroom.

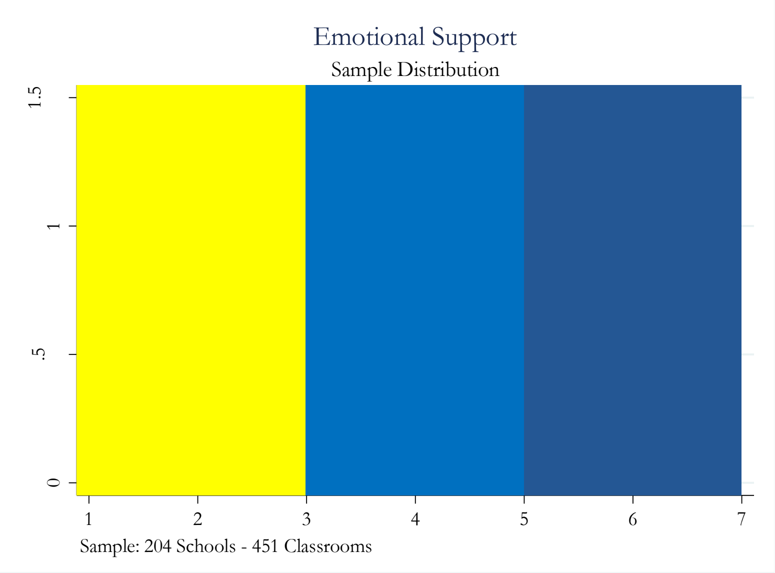
What is the CLASS?

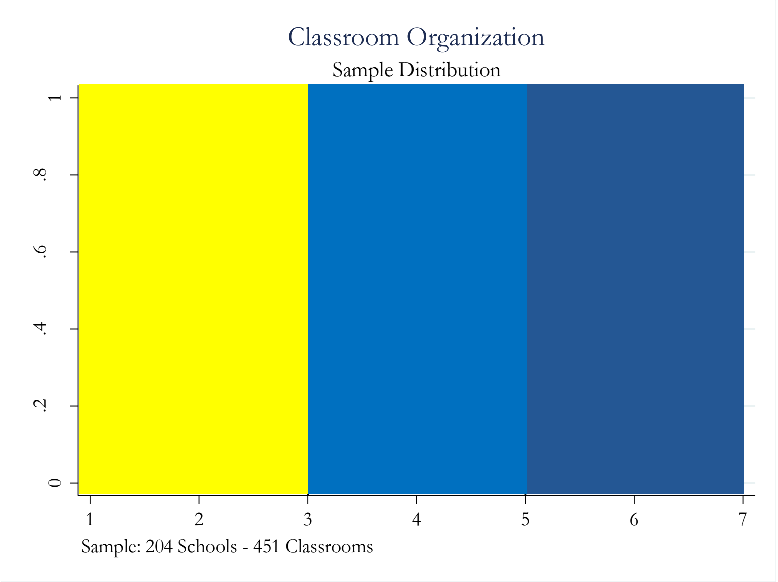
- The CLASS has been adopted by Head Start as part of its monitoring process, and several states have also adopted the CLASS as part of their Quality Rating and Improvement Systems (Hamre et al. 2012)
- Numerous papers in the US have shown positive correlations between CLASS scores and child learning outcomes (for example, Mashburn et al. 2008, and cites therein)
- A number of randomized trials in US have shown that the CLASS is malleable
  - Pianta et al. (2008) and Downer et al. (2011) on a web-based teacher professional development program known as My Teaching Partner
  - Hamre et al. (2012) and Brown et al. (2010) on training and coaching programs
  - Domitrovich et al. (2008) on an enriched version of Head Start
- A handful of randomized trials in the US has also shown that some interventions that improve CLASS scores also result in better child learning outcomes, although the results are mixed (Downer et al. 2011, 2013; Bierman et al. 2008)
- In Chile, Yoshikawa et al. (2013) find that a teacher training and mentoring program (*Un Buen Comienzo*) has short-term effects on the CLASS but no discernible effects on child outcomes

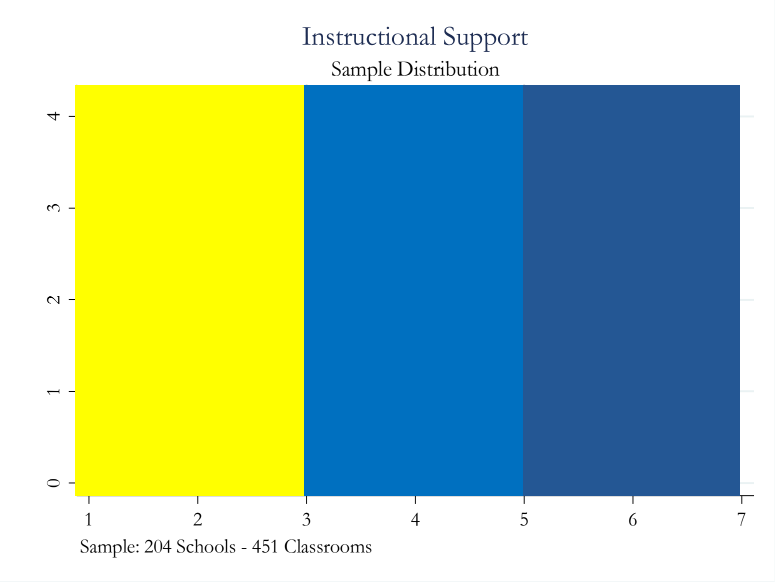
Application of CLASS in Ecuador

- 451 teachers in sample filmed for full day during 2012/13 school year
- Of these 451 teachers, 342 had also been filmed in the 2011/12 school year (44 teachers filmed for more than one day)
- Discard first hour, and discard lunch, break, and time spent on other "non-educational" activities, as well as time without the main teacher
- The rest of the video is cut into 20-minute segments
- 4 segments chosen at random, and assigned randomly to 2 coders
  - If difference in scores for "high-variation" dimensions (positive climate, teacher sensitivity, regard for student perspectives, behavior management, productivity, instructional learning formats) is larger than 2 points, segment sent to third coder
  - If difference in scores for "low-variation" dimensions (negative climate, concept development, quality of feedback, language modeling) is larger than 1 point, segment sent to third coder
  - Keep scores from the two coders whose scores are most similar to each other
  - In practice, only 125 segments (3.4% of all segments) were sent for a third coding in 2012/13

CLASS scores in Ecuador

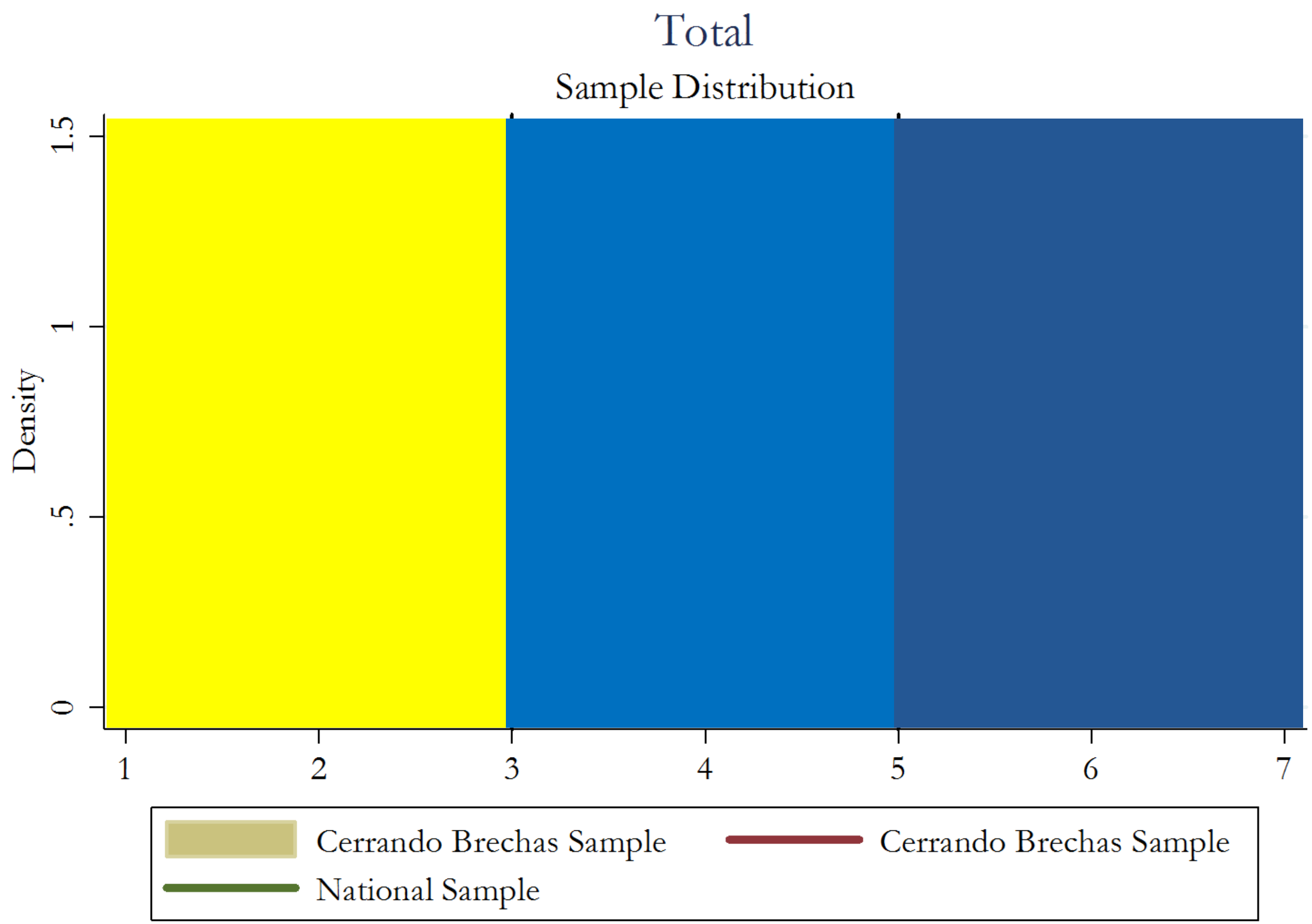






	Score variation	
	Between Schools	Within Schools
Emotional Support	0.41	0.59
Classroom Organization	0.38	0.62
Instructional Support	0.44	0.56

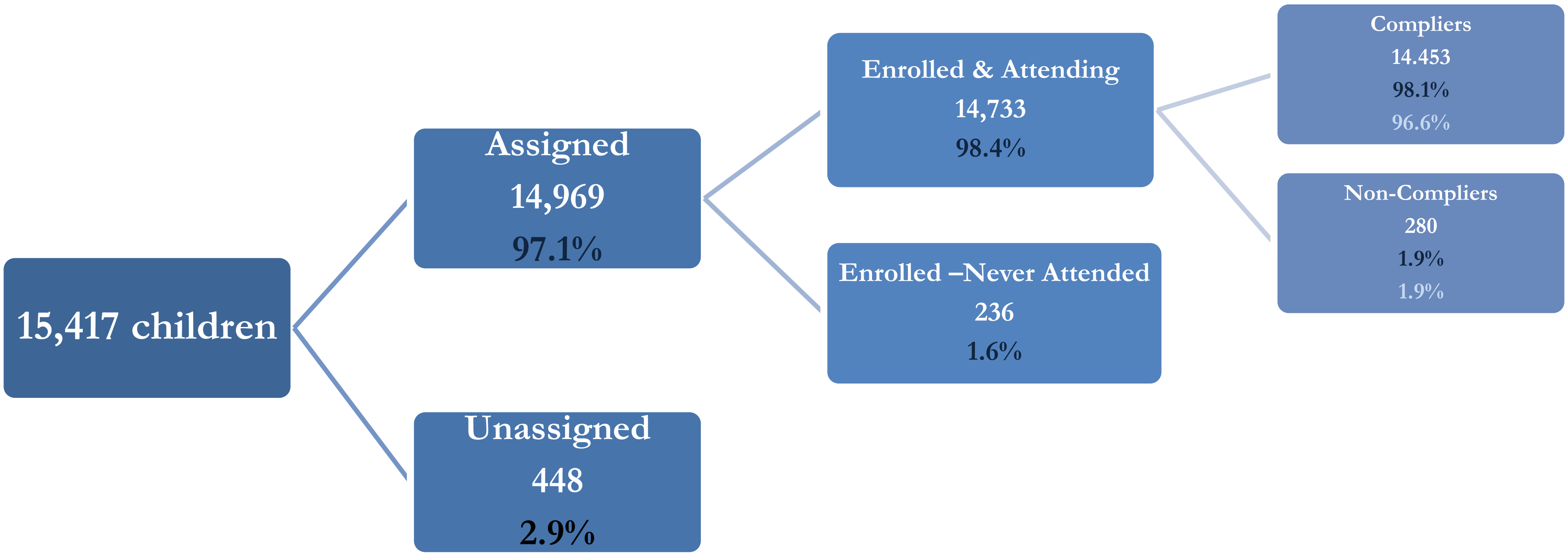
CLASS scores in Ecuador



CLASS scores in Ecuador

Table 6: Measurement error in CLASS				
	Mean score (1-7)	Standard deviation	Reliability ratio	
			Coders	Segments
Emotional support	4.07	0.33	0.89	0.59
Classroom organization	4.79	0.47	0.89	0.57
Instructional support	1.15	0.18	0.85	0.29
Total CLASS score	3.41	0.28	**0.92**	**0.61**
Note: The reliability ratio is given by Cov(Score1, Score2)/Variance(Avg. Score)				

Did the experiment "work"?



Did the experiment “work”?

- Check for associations between child baseline characteristics and (previous year) teacher characteristics
- 52 teachers (12 percent) left our study schools between 2011/12 (the year before our student results) and 2012/13, and 86 teachers (19 percent) left these schools during the course of the 2012/13 school year
  - Who are the teachers who leave?
  - Is it more likely that a teacher will leave if she is assigned a more difficult set of students (for example, lower TVIP students, younger students)
- Check for differences between assigned and unassigned children, children who dropped out and those who did not, children missing TVIP or other baseline covariates and other children missing end-of-year test data and other children (Appendix table)
- Non-compliers assigned to classes they were *assigned* to, rather than those they actually *attend* (ITT analysis)

Associations between child and teacher characteristics at baseline

Table 7: Associations between child and teacher characteristics at baseline

	Mean	CLASS 2011/12	Years of Experience
Age, months	60.39 (5.08)	-0.002 (0.071) [12,000]	-0.005 (0.008) [12,034]
Female	0.49 (0.50)	0.01 (0.007) [12,055]	-0.001 (0.001) [12,089]
TVIP	0.00 (1.00)	-0.025 (0.017) [11,289]	0.003 (0.002) [11,321]
Wealth index	0.00 (0.99)	0.002 (0.016) [11,305]	0.0006 (0.002) [11,336]
Mother's education	8.79 (3.8)	0.087 (0.052) [10,717]	-0.002 (0.007) [10,744]

Notes: Coefficient, standard error (in parentheses) and sample size (in square brackets). Dependent variable is given in first column, explanatory variables (CLASS, years of experience) in first row of table. All regressions include school fixed effects. Standard errors clustered at school level.

Did the experiment “work”?

.

Table 8: Characteristics of teachers who left sample					
		Between years		Within year	
	Mean	w/out FE (n=453)	With FE (n=116)	w/out FE (n=400)	With FE (n=107)
CLASS 2011/12	3.65	-0.05 (0.06)	0.06 (0.09)	-0.10 (0.05)*	-0.07 (0.09)
Total years of experience	15.11	-0.31 (1.47)	1.56 (2.12)	-6.38 (1.0)***	-4.09 (2.24)*
Years teaching in current school	8.37	-1.67 (1.32)	-0.41 (1.71)	-4.84 (0.65)***	-4.0 (1.67)**
Tenured	0.61	-0.25 (0.07) ***		-0.44 (0.05)***	

Notes: Sample is all teachers in sample schools at the beginning of 2012/13 school year. Dependent variable is a dummy for teachers who left sample, explanatory variable is given in first column of table. Standard errors clustered at school level. 54 teachers left the sample between the 2011/12 and 2012/13 school years and 74 within the 2012/13 school year.
** significant at 5% level, ***at 1% level.

Did the experiment "work"?

Table 9: Characteristics of students in classes where teacher left school within year			
	Mean	w/out FE	With FE
TVIP	0.02	-0.12 (0.04)*** [13,479]	-0.03 (0.03) [3,297]
Age	60.34	-0.29 (0.14)** [13.467]	-0.06 (0.13) [3,292]
Girl	0.49	0.002 (0.01) [13,479]	0.013 (0.01) [3,297]
Mother's Education	8.81	-0.18 (0.17) [12,714]	-0.04 (0.09) [3,117]
Wealth	0.01	-0.08 (0.06) [13,428]	-0.003 (0.03) [3,282]

Notes: Sample is all teachers in sample schools at the beginning of 2012/13 school year. Dependent variable is a dummy for teachers who left sample, explanatory variable is given in first column of table. Standard errors clustered at school level. 52 teachers left the sample between the 2011/12 and 2012/13 school years and 74 within the 2012/13 school year. ** significant at 5% level, ***at 1% level.

Did the experiment "work"?

Conclusions:

- 97 percent of children were assigned and, of those who were assigned, 98 percent comply with the randomization protocol
- Assigned and unassigned children, and compliers and non-compliers, are similar in their observable characteristics and in those of their teachers: There is no evidence of purposeful (endogenous) matching of teachers and students and, given the very few children involved, the potential for bias would appear to be very limited
- There are no observable differences in the characteristics of students assigned to better and worse teachers (lagged CLASS score, experience)
- 11 percent of teachers leave the school they are teaching at between the end of one year and the beginning of the next year, and 19 percent leave during the course of the 2012/13 school year
  - Teachers who exit the sample are less likely to be tenured, and they have fewer years of experience
  - There is no evidence that teachers who get a "bad" draw of children (in spite of the random assignment) are more likely to leave the school in that year: No obvious concerns of internal validity

Study design

1. **Random assignment sample**: Includes all children who nominally enrolled at the beginning of the 2012/13 school year and their teachers: 204 schools, 451 teachers, 15,417 children
   – This is the sample with which we conducted the random assignment
2. **Classroom effects sample**: Includes all children who attended school at the beginning of the 2012/13 school year, were enrolled at the end of the school year, and have end-of-year test results: 204 schools, 451 teachers, 13,484 children
   - This is the sample with which we analyze classroom effects
3. **Full teacher effects sample**: Classroom effects sample but limited to observations in which teacher was the same at the beginning and end of 2012/13 school year, and data on teacher characteristics and 2012/13 CLASS are available: 149 schools, 332 teachers, 9,821 children
   – This is the sample we use to run OLS regressions of child learning outcomes on teacher characteristics
   – 86 teachers moved within 2012/13 school year
   – Note that because identification comes from within-school, cross-classroom variation, when we lose a teacher in a two-classroom school, we lose all of the observations for that school
4. **Restricted teacher sample**: Full teacher sample but restricted to teachers for whom we also have CLASS from 2011/12 school year: 121 schools, 258 teachers, 7,582 children
   – This is the sample we use to run the IV regressions
   – 54 teachers moved between 2011/12 and 2012/13 school years

Outline

- Motivation
- Some basic facts about Ecuador
- Identification challenges and study design
- Descriptive results: The CLASS
- Classroom effects
- Teacher characteristics, behaviors, and student learning outcomes
- Extensions and work currently under way

Classroom effects: methodological discussion

First approach

- Step 1: Residualize test scores
  - Regress end-of-year test score of child i with teacher t in school s on her age, gender, baseline TVIP, mother education and a wealth aggregate
  - Predict residual
- Step 2: In each class, calculate the average residualized score for each test, and calculate the averages across tests in language, math, and executive function
- Step 3: Calculate the average difference across classrooms within the same school, and the associated standard errors
  - When there are more than 2 classrooms, keep the "best" and "worst"
- Step 4: Correct for measurement error
- Step 5: See whether in classrooms where there is more learning in one dimension there is also more learning in another dimension (correlation matrix)

- **A point of interpretation**: Note that these classroom effects include the effect of teachers as well as the effects of any differences in peers across classrooms (in spite of random assignment), or any other differences between one and another classroom

Classroom effects: methodological discussion

Why are we concerned about measurement error?

- Intuition: With measurement error, two teachers who have identical *true* effects will have different *observed* effects
- Because we are taking the absolute value of the difference between teachers, the measurement error does not "cancel out"—at "low" values of the true difference, measurement error will *increase* the observed difference

Solution:

- For each pair of teachers, we can calculate p, the probability that the observed difference $\delta \geq 0$, and $(1 - p_s)$, the probability that the observed difference $\delta < 0$.
- Then, calculate the corrected difference for each school:

$$\delta_s^C = \delta_s^o - (1 - p_s)\delta_s^o$$

- Finally, take the average of these corrected differences

Classroom effects on learning

		Uncorrected	Corrected	Corrected & Weighted
Language & early literacy	Letter & word ID	0.307	0.213	0.208
	ID of first sound of words	0.338	0.263	0.257
	Receptive vocabulary	0.286	0.194	0.190
	Oral comprehension	0.270	0.180	0.182
	Language Total	**0.340**	**0.260**	**0.258**
Math	Number ID	0.322	0.235	0.235
	Block rotation	0.248	0.150	0.147
	Numeric series	0.295	0.215	0.214
	Applied problems	0.278	0.181	0.177
	Math Total	**0.317**	**0.222**	**0.220**
Executive Function	Inhibitory control	0.233	0.142	0.139
	Attention control	0.269	0.170	0.164
	Cognitive flexibility	0.261	0.167	0.160
	Working memory	0.227	0.127	0.124
	Executive Function	**0.263**	**0.167**	**0.159**
Total		**0.326**	**0.244**	**0.240**

Note: Calculations based on the classroom effects sample. The values reported in the third and fourth columns are the difference in learning outcomes between teachers in a given school. When there is more than one teacher, we calculate the difference between the "best" teacher in a school (the teacher who has the largest impact on learning) and "worst" teacher (the teacher who has the smallest impact on learning)

Classroom effects on learning

		Language					Math					Executive Function					Total
		Letter Word ID	Sound ID	Vocabulary (TVIP)	Oral comprehension	**Language (Total)**	Number ID	Block Rotation	Number Series	Applied Problems	**Math (Total)**	Response Inhibition	Attention Control	Cognitive Flexibility	Working Memory	**EF (Total)**	**Total**
Language	Letter Word ID	1															
	Sound ID	0.53	1														
	Vocabulary (TVIP)	0.32	0.32	1													
	Oral comprehension	0.17	0.20	0.46	1												
	Language (Total)	0.75	0.81	0.66	0.57	1											
Math	Number ID	0.60	0.41	0.44	0.30	0.62	1										
	Block Rotation	0.22	0.20	0.32	0.27	0.34	0.30	1									
	Number Series	0.50	0.45	0.50	0.34	0.63	0.73	0.23	1								
	Applied Problems	0.46	0.40	0.48	0.37	0.59	0.44	0.34	0.48	1							
	Math (Total)	0.60	0.49	0.58	0.43	0.73	0.84	0.59	0.83	0.74	1						
Executive Function	Response Inhibition	0.14	0.10	0.19	0.10	0.18	0.26	0.30	0.12	0.14	0.27	1					
	Attention Control	0.28	0.19	0.39	0.35	0.40	0.35	0.39	0.33	0.34	0.47	0.22	1				
	Cognitive Flexibility	0.24	0.13	0.16	0.17	0.24	0.29	0.22	0.29	0.27	0.36	0.08	0.30	1			
	Working Memory	0.21	0.29	0.30	0.25	0.37	0.27	0.17	0.24	0.21	0.30	-0.01	0.28	0.22	1		
	EF (Total)	0.35	0.28	0.41	0.35	0.47	0.47	0.44	0.39	0.38	0.56	0.54	0.72	0.66	0.57	1	
Total		0.68	0.65	0.66	0.54	0.88	0.77	0.53	0.74	0.69	0.91	0.36	0.59	0.46	0.46	0.75	1

N = 204

Classroom effects: methodological discussion

Second approach

- Step 1: Calculate (uncorrected) teacher effects
  - Regress end-of-year test score of child i with teacher t in school s on her age, gender, baseline TVIP, mother's education, a wealth aggregate, and teacher FEs.
- Step 2: Correct for measurement error

  Note that:

$$\tau_t^o = \tau_t^* + e_t$$
$$Var(\tau_t^*) = Var(\tau_t^o) - Var(e_t)$$

- Empirical Bayes correction (as in Kane and Staiger 2002; Jacob and Lefgren 2008)
- For each teacher effect, calculate the standard error of difference and then, square each one to get the variance of the difference:
  - Take average: this is an estimate of $Var(e_t)$.
- Calculate $Var(\tau_t^o)$ as the variance of the estimated teacher fixed effects.
- With this in hand, calculate $Var(\tau_t^*)$
- **A point of interpretation**: Note that these teacher effects are *not* identified exclusively of the random assignment, they include both within-school, cross-classroom comparisons, as well as cross-school comparisons (~Chetty et al. 2013)

Classroom effects on learning: Magnitude

- The measurement error-corrected estimates indicate that, in Ecuador, a one-standard deviation improvement in teacher quality raises student outcomes by 0.31 standard deviations in reading, 0.27 standard deviations in math, and 0.24 standard deviations in executive function
- These are large effects relative to those found in the US:
  - Hanushek and Rivkin (2012) summarize the findings from 10 studies of teacher effects
    - Across all studies, a one-standard deviation increase in teacher quality raises math outcomes by 0.17 standard deviations, and reading outcomes by 0.13 standard deviations, on average
  - More recent work by Chetty et al. (2013) estimates that a one-standard deviation better teacher improves child outcomes by 0.14 standard deviatoions in math, and 0.10 standard deviations in English

Outline

- Motivation
- Some basic facts about Ecuador
- Identification challenges and study design
- Descriptive results: The CLASS
- Classroom effects
- Teacher characteristics, behaviors, and student learning outcomes
- Extensions and work currently under way

Methodological aside

Basic approach:

$Y_{its} = \alpha_s + \beta_1 X_i + \beta_2 CLASS_t + \beta_3 X_t + \varepsilon_{its}$, k=1, 2… 12

- Note that this gives results that are very similar to those from a 3-step approach, with the weights given by the number of students per classroom
  - Step 1: Residualize scores, as before
  - Step 2: Calculate classroom averages of scores, CLASS
  - Step 3: Run regressions of the averages

$Y_{ts} = \alpha_s + \beta_2 CLASS_t + \beta_3 X_t + \varepsilon_{ts}$, k=1, 2… 12

  - In general, the one-step approach will be more efficient

Methodological aside

<u>First challenge: correcting for measurement error in CLASS</u>

Two concerns about measurement error:

- <u>Concern 1</u>: We are using contemporaneous CLASS and learning outcomes so there could be positively-correlated shocks
  - In general, this will lead the coefficient β_2 to be upward-biased
- <u>Concern 2</u>: “Classical” measurement error
  - Coder error
  - Differences in scores across segments within days, across days within school year, across school years
  - This suggests there is likely to be substantial attenuation bias
- Solution: Instrumental Variables estimation (Imbens and Angrist 1994): Instrument 2012/13 CLASS with 2011/12 CLASS

Methodological aside

<u>Second challenge: getting the "right" standard errors</u>

- Two separate issues
- Issue 1: Need to account for the fact that we are testing 12 different hypotheses (36 if we consider the three different dimensions of the CLASS separately)
- Two approaches in the literature:
  - <u>Approach 1</u>: Do no test individual outcomes, rather test families of outcomes (Kling, Liebman and Katz 2007): Aggregate 4 language and early literacy tests, 4 math tests, 4 EF tests, as well as all 12 tests
  - <u>Approach 2</u>: Step-down procedure (Romano and Wolf 2005): Iterative bootstrapping procedure, results in revised t-statistics for rejecting null at 1%, 5%, 10%
    - In our case, the revised t-statistic is 3.38 to reject the null with 95% confidence
- Issue 2: need to allow for spatially-correlated standard errors (Moulton 1986)
  - Main results: allow for clustering at school level
  - Robustness check: allow for clustering at classroom level

Methodological aside

<u>Romano & Wolf</u> Step-down estimation (Econometrica, 2005)

Getting started:

1. There is total of R regressions of children outcomes (C) on teacher characteristics (T), where $R = C \times T$
   – We have 12 outcomes (4 math tests, 4 language tests, 4 EF tests), and 1 explanatory variable (CLASS)
2. Store all of the t statistics from the regressions: a total of R **original $\boldsymbol{t}$**s.
   – We have 12 original *ts*

Bootstrap procedure:

1. Draw I random samples of classrooms with replacement. The number of classrooms in each sample will be the same as in the original sample.
   – We draw 1,000 samples of 149 classrooms
2. Estimate the set of R regressions of child outcomes on teacher characteristics in each new sample, and store the t-statistic ($R \times I$ **bootstrap $\boldsymbol{t}$**s)
   – In each of the 1,000 samples, run the 12 regressions, and store 12 *ts* (for a total of 12,000 bootstrap *ts)*

Methodological aside

Romano-Wolf algorithm:

Step 1:

1. From each of the I sets of estimations, obtain the highest t-statistic ($\boldsymbol{t_1}$ **max**) from the R number of t-statistics.
   – In each of the 1,000 sets of estimations, store the *highest* of the 12 *t*-statistics
2. From the distribution of $\boldsymbol{t_1}$ **max** (a total of I), set the α^{th} (90^{th}, 95^{th} or 99^{th}) percentile as the first critical value ($\boldsymbol{t_{c1}}$).
   – We now have a distribution of 1,000 $\boldsymbol{t_1}$ **max** statistics: Choose the 90^{th} (the 900^{th} highest *t*), the 95^{th} (the 950^{th} highest *t*), and the 99^{th} (the 990^{th} highest *t*). In our case, these values are 2.74, 3.01 and 3.52 respectively
3. Compare each of the R **original** $\boldsymbol{t}$s with the critical value $\boldsymbol{t_{c1}}$. Coefficients for which **original** $|\boldsymbol{t}| > \boldsymbol{t_{c1}}$ will be significant at the α^{th} confidence level.
   – In our case, 7 (at the 90^{th} confidence level), 6 (at the 95^{th}), and 4 (at the 99^{th}) of the original t-statistics have a value that is higher than the critical value.

Methodological aside

<u>Step 2</u>:

1. Exclude the S_1 **bootstrap** $\boldsymbol{t}$s corresponding to the significant coefficients on Step 1.
   - We excluded 7,000 (90^{th}), 6,000 (95^{th}), and 4,000 (99^{th}) *t*s out of the original set of 12,000 *t*s, and we have a remaining 5,000 (90^{th}), 6,000 (95^{th}), and 8,000 (99^{th}) *t*s
2. With the remaining $(R - S_1)$ **bootstrap** $\boldsymbol{t}$s, obtain the new highest *t*-statistic ($\boldsymbol{t_2}$ **max**) in each of the I sets of estimations.
   - In each of the 1,000 sets of estimations, store the highest of the 5 remaining *t*s (at the 90^{th} level), 6 remaining *t*s (95^{th}), or 8 remaining *t*s (99^{th})
3. From the distribution of $\boldsymbol{t_2}$ **max** (a total of I), set the α^{th} (90^{th}, 95^{th} or 99^{th}) percentile as the second critical value ($\boldsymbol{t_{c2}}$).
   - We now have a distribution of 1,000 $\boldsymbol{t_2}$ **max** statistics: Choose the 90^{th} (the 900^{th} highest *t*), the 95^{th} (the 950^{th} highest *t*), and the 99^{th} (the 990^{th} highest *t*). These values are 2.42, 2.75 and 3.49 respectively
4. Compare each of the remaining $R - S_1$ **original** $\boldsymbol{t}$s with the critical value $\boldsymbol{t_{c2}}$. Coefficients for which **original** $|\boldsymbol{t}|> \boldsymbol{t_{c2}}$ will be significant at the α^{th} confidence level.
   - In our case, 0 (90^{th}), 1 (95^{th}), and 0 (99^{th}), of the original *t*-statistics have a value that is higher than the critical value $\boldsymbol{t_{c2}}$.

- The algorithm will continue until no coefficient is significant or all of the coefficients are significant on a given step.
  - We have a total of 2 steps for the 90^{th} and 99^{th} confidence levels and 3 for the 99^{th}.

CLASS results: OLS, individual tests

		(i)	(ii)
Language & early literacy	Letter & word ID	0.06 (0.017)***	0.05 (0.019)*
	ID of first sound of words	0.08 (0.025)***	0.08 (0.026)**
	Receptive vocabulary	0.04 (0.012)***	0.04 (0.012)**
	Oral comprehension	0.02 (0.013)	0.02 (0.014)
Math	Number ID	0.08 (0.018)***	0.08 (0.019)***
	Block rotation	0.03 (0.016)	0.03 (0.016)
	Numeric series	0.06 (0.015)***	0.05 (0.016)**
	Applied problems	0.05 (0.015)***	0.05 (0.015)**
Executive function	Inhibitory control	0.02 (0.017)	0.02 (0.016)
	Attention control	0.03 (0.012)**	0.03 (0.013)*
	Cognitive flexibility	0.03 (0.016)	0.03 (0.016)
	Working memory	0.01 (0.016)	0.01 (0.017)

Note: The table reports the coefficient on the CLASS score and the standard error, adjusted for clustering at the school level. All regressions include the baseline child and household characteristics and school fixed effects; specification (ii) also includes teacher experience and tenure status. We also test whether each coefficient is statistically different from zero after adjusting the critical values for multiple hypothesis testing (Romano and Wolf, 2005): ** significant at 5%, *** at 1%

CLASS results: OLS, grouped tests

	Language	Math	Executive Function	TOTAL
	OLS: Full teacher effects sample			
CLASS (Total)	0.06	0.07	0.04	0.07
	(0.017)***	(0.016)***	(0.014)**	(0.016)***
Experience (<=3 years)	-0.12	-0.07	-0.07	-0.10
	(0.056)**	(0.06)	(0.035)**	(0.046)**
Tenured	0.03	0.04	-0.03	0.02
	(0.038)	(0.032)	(0.033)	(0.034)

Note: The table reports the coefficient on the CLASS score, experience and tenure and the standard error, adjusted for clustering at the school level. All regressions include the baseline child and household characteristics and school fixed effects ** significant at 5%, *** at 1%
The CLASS explains 9.8 percent of the within-school, cross-classroom variation in test scores. When experience enters as a single (linear) term, it explains 1.1 percent of the variation in test scores; when it enters as single dummies for every year of experience, it explains 6.8 percent of the variation in test scores; when it enters as a dummy for teachers with <-3 years of experience, it explains 2.9 percent of the variation in test scores. The dummy variable for tenured teachers explains 1.1 percent of the within-school, cross-teacher variation in scores.

CLASS results: IV, grouped tests

	Language	Math	Executive Function	TOTAL
	OLS: Restricted teacher effects sample			
CLASS (Total)	0.04	0.05	0.02	0.05
	(0.02)**	(0.02)***	(0.017)	(0.019)**
Experience	-0.05	-0.06	-0.05	-0.06
	(0.069)	(0.106)	(0.05)	(0.067)
Tenured	0.04	0.06	-0.02	0.03
	(0.046)	(0.04)	(0.041)	(0.041)
	Instrumental variables			
CLASS (Total)	0.14	0.18	0.16	0.18
	(0.078)	(0.083)**	(0.073)**	(0.078)**
Experience	-0.03	-0.04	-0.02	-0.03
	(0.061)	(0.094)	(0.061)	(0.057)
Tenured	0.01	0.02	-0.07	-0.01
	(0.047)	(0.048)	(0.046)	(0.046)

Note: The table reports the coefficient on the CLASS score, experience and tenure and the standard error, adjusted for clustering at the school level. All regressions include the baseline child and household characteristics and school fixed effects. The coefficient on the first stage of the instrumental variables regression is 0.31 (0.071). ** significant at 5%, *** at 1%

CLASS results: robustness checks

	Executive Function	Language	Math	Total
Clustered at school level	0.04	0.06	0.07	0.07
	(0.014)**	(0.017)***	(0.016)***	(0.016)***
Clustered at classroom level	0.04	0.06	0.07	0.07
	(0.011)***	(0.012)***	(0.012)***	(0.012)***
Three-step approach	0.03	0.06	0.07	0.06
	(0.019)	(0.022)***	(0.022)***	(0.021)***
Excluding children who were unassigned	0.03	0.06	0.07	0.06
	(0.015)**	(0.017)***	(0.017)***	(0.016)***
Excluding schools where there was at least one violation of random assignment	0.03	0.07	0.07	0.07
	(0.017)	(0.018)***	(0.016)***	(0.017)***
CLASS constructed by principal components	0.04	0.07	0.07	0.07
	(0.014)***	(0.017)***	(0.016)***	(0.016)***

Note: The table reports the coefficient on the CLASS score and the standard error All regressions include the baseline child and household characteristics and school fixed effects ** significant at 5%, *** at 1%

Heterogeneity: Child characteristics

	Executive Function	Language	Math	All Tests
CLASS	0.036	0.072	0.075	0.072
	(0.014)**	(0.016)***	(0.016)***	(0.015)***
CLASS * TVIP	-0.010	0.009	-0.010	-0.004
	(0.01)	(0.012)	(0.009)	(0.01)
TVIP	0.425	0.554	0.458	0.559
	(0.012)***	(0.012)***	(0.011)***	(0.011)***
CLASS	0.051	0.068	0.069	0.073
	(0.02)**	(0.023)***	(0.02)***	(0.022)***
CLASS * Female	-0.030	0.011	0.014	0.000
	(0.018)*	(0.019)	(0.017)	(0.017)
Female	-0.001	0.007	-0.066	-0.025
	(0.018)	(0.018)	(0.018)***	(0.018)
CLASS	0.037	0.075	0.078	0.075
	(0.017)**	(0.02)***	(0.017)***	(0.019)***
CLASS * AGE	0.000	0.001	0.003	0.001
	(0.003)	(0.002)	(0.003)	(0.003)
AGE	0.020	0.023	0.027	0.028
	(0.003)***	(0.003)***	(0.003)***	(0.003)***

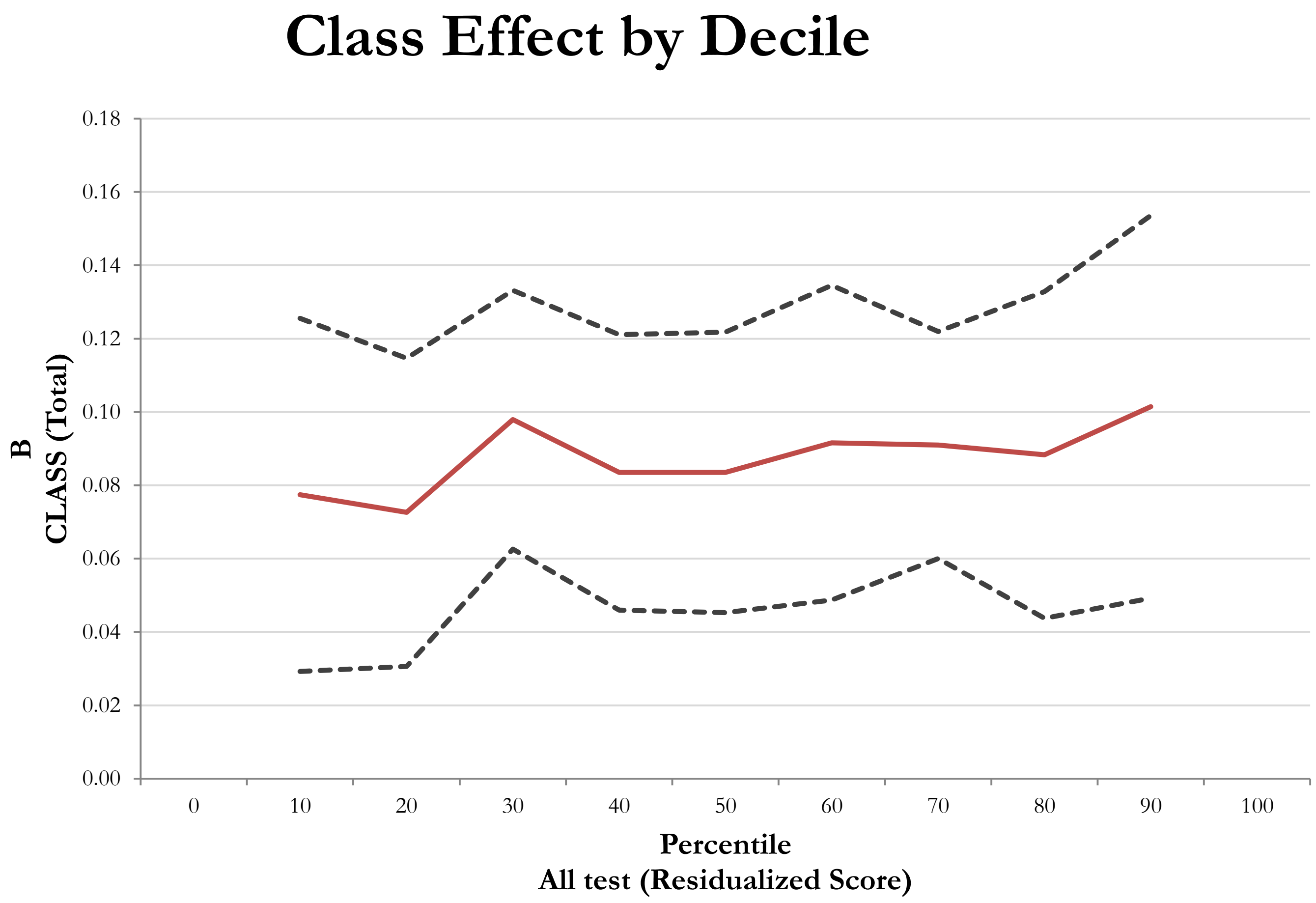
Note: All regressions include the baseline child and household characteristics and school fixed effects . Standard errors clustered at the school level. ** significant at 5%, *** at 1%

Heterogeneity: family characteristics

	Executive Function	Language	Math	All Tests
CLASS	0.036	0.072	0.076	0.073
	(0.016)**	(0.02)***	(0.017)***	(0.018)***
CLASS * Mother's Education	0.001	0.001	-0.002	0.000
	(0.003)	(0.003)	(0.003)	(0.003)
Mother's Education	0.037	0.054	0.043	0.052
	(0.002)***	(0.003)***	(0.003)***	(0.002)***
CLASS	0.036	0.072	0.075	0.072
	(0.017)**	(0.02)***	(0.017)***	(0.018)***
CLASS * Wealth Index	-0.012	0.001	-0.003	-0.005
	(0.012)	(0.012)	(0.009)	(0.011)
Wealth Index	0.090	0.135	0.094	0.124
	(0.012)***	(0.013)***	(0.01)***	(0.011)***

Note: All regressions include the baseline child and household characteristics and school fixed effects . Standard errors clustered at the school level. ** significant at 5%, *** at 1%.

Heterogeneity: quantile regressions



Teacher quality: Extensive and intensive margins

	Number of days absent	
	(I)	(II)
CLASS (Total)	-0.082	-0.072
	(0.131)	(0.13)
Experience	0.288	0.264
	(0.373)	(0.376)
Tenured	0.008	0.014
	(0.269)	(0.267)

Note: All regression include school fixed effects; specification (ii) also includes the baseline child and household characteristics. Dependent variable is days absent between April and July. Mean: 2.8, Median: 2.0, 10th percentile 0, 90th percentile 8. All regressions include school fixed effects. Standard errors clustered at school level. Significant * at 10%, ** at 5%, *** at 1%.

It is conceivable that children assigned to better teachers miss fewer days (the "extensive" margin), in addition to learning more for any given level of attendance

However, that does not appear to be the case

Can parents recognize better teachers?

	OLS	Ordered Probit	OLS Dummy
CLASS	0.067	0.146	0.055
	(0.014)***	(0.03)***	(0.011)***
	[9706]	[9706]	[9706]
Average Score (-i)	0.179	0.383	0.134
	(0.057)***	(0.12)***	(0.043)***
	[9706]	[9706]	[9706]
Experience (<=3 years)	-0.154	-0.317	-0.128
	(0.044)***	(0.088)***	(0.039)***
	[9706]	[9706]	[9706]
Tenured	0.087	0.184	0.07
	(0.028)***	(0.057)***	(0.022)***
	[9706]	[9706]	[9706]

Notes: Regressions of parent perception of teacher quality (1-5 scale, 1 = very bad, 5 = very good) on measures of teacher quality. In OLS Dummy, the dummy variable takes the value of one if parent's perception of the teacher is very good , zero otherwise. Average parents' perception is 4.5, the median is 5.0; 0.1% of teachers are classified as very bad, 0.2% a bad, 4.2% as average, 37.1% as good, and 58.4% as very good.
All regressions include school fixed effects. Standard errors clustered at school level.

Do parents *respond* to differences in teacher quality?

Dummy Without baseline controls	Mean (SD)		CLASS (TOTAL)		Test Result: Classroom Average (-i)	
			I	II	I	II
Read books, watch pictures or drawings in a book with child *N* = 9,782	1.22	**β**	0.02	0.01	0.09	0.03
	(1.89)	**s.e.**	(0.028)	(0.008)	(0.101)	(0.028)
Tell short stories or tales to child *N* = 9,784	1.13	**β**	0.03	0.01	0.06	0.01
	(1.79)	**s.e.**	(0.031)	(0.009)	(0.099)	(0.029)
Sing to child or sing with, even lullabies *N* = 9,786	2.61	**β**	0.02	0.00	-0.38	-0.04
	(2.64)	**s.e.**	(0.037)	(0.006)	(0.149)**	(0.023)
Take child outside. Go to the park or go for a walk *N* = 9,794	1.71	**β**	0.01	0.00	-0.01	-0.03
	(1.7)	**s.e.**	(0.026)	(0.006)	(0.086)	(0.022)
Play with child with toys *N* = 9,780	1.81	**β**	0.04	0.01	0.07	0.00
	(2.38)	**s.e.**	(0.038)	(0.008)	(0.135)	(0.031)
Drawing or painting with child *N* = 9,790	2.59	**β**	0.02	0.01	0.05	0.01
	(2.55)	**s.e.**	(0.045)	(0.007)	(0.149)	(0.028)
Play with child to name or count objects or colors *N* = 9,786	3.2	**β**	0.06	0.01	-0.15	-0.02
	(2.58)	**s.e.**	(0.043)	(0.006)*	(0.182)	(0.028)

Notes: Regression of teacher quality on: (I) number of times parents do each activity with the child during the past seven days, and (II) on a dummy variable that the value of one if parents did each activity with the child during the past seven days. All regressions include school fixed effects. Standard errors clustered at school level.

Conclusions

1. **<u>There are substantial classroom effects</u>**:
   - After accounting for initial differences across children, and correcting for measurement error, the average end-of-year difference in learning outcomes for children randomly assigned to one or another teacher are 0.26 sds for language, 0.22 sds for math, and 0.17 sds for EF
   - In a measure more directly comparable to the US literature: A one standard deviation better teacher increases student learning by 0.31 sds in language, 0.27 sds for math, and 0.24 sds for executive function
   - Teacher effects appear to be substantially larger than in the US

- Being assigned to a better teacher closes approximately 1/2 the gap in learning outcomes between children of mothers with complete primary education and those of mothers with complete secondary education—in a single year
- In classrooms where there is more learning in one dimension (say, math) there is also more learning in other dimensions (say, executive function): It does not appear that teachers "specialize", some are better at everything

Conclusions

2. <u>Teacher characteristics and (especially) behaviors matter:</u>

- Children randomly assigned to teachers who provide better socio-emotional support, have better classroom management skills and provide better instructional support have significantly better learning outcomes
  - A 1 sd increase in the CLASS is associated with .07 sd increase in child learning outcomes (OLS), and .18 sd increase after correcting for measurement error (IV)
  - Within-school, cross-classroom differences in the CLASS explain approximately 10 percent of the within-school, cross-classroom differences in learning outcomes
- Children randomly assigned to teachers with 3 or fewer years of experience learn 0.10 sds less than those assigned to more experienced teachers
  - After ~the third year of experience, the experience learning profile flattens out (much as in the US)
- Teachers who are tenured are no more effective, on average, than those who work on a contract basis (much as in the US, see Kane et al. 2008)
- Parents recognize better teachers, but do not adjust their behaviors to take account of these differences in teacher quality (as opposed to what is reported by Pop-Eleches and Urquiola 2013 for Romania)

Conclusions

- Krueger (1999): “One well-designed experiment should trump a phalanx of poorly-controlled, imprecise observational studies based on uncertain statistical specifications”.

Annex 1: References

- Almond, D., and J. Currie. 2011. "Human Capital Development before Age Five." In O. Ashenfelter and D. Card, Eds., *Handbook of Labor Economics*, Volume 4, Elsevier.
- Andrabi, T., J. Das, A. I. Khwaja, and Tr. Zajonc. 2011. "Do Value-Added Estimates Add Value? Accounting for Learning Dynamics." *American Economic Journal: Applied Economics* 3(3): 29-54.
- Attanasio, O., S. Grantham-McGregor, C. Fernandez, E. Fitzsimons, M. Rubio-Codina, and C. Meghir. 2013. "Enriching the Home Environment of Low-Income Families in Colombia: A Strategy to Promote Child Development at Scale." Unpublished.
- Barros, Ricardo Paes et al. 2013. "Does Public Daycare Help Low-Income Families and their Children? Experimental Evidence from Rio de Janeiro City." Unpublished.
- Bastos, P., N. Bottan and J. Cristia. 2012. "Access to Preprimary Education and Progression in Primary School: Evidence from Rural Guatemala." IDB Working Paper 4815.
- Bierman, K., C. Domitrovich, R. Nix, S. Gest, J. Welsh, M. Greenberg, C. Blair, K. Nelson, and S. Gill. 2008. "Promoting Academic and Social-Emotional School Readiness: The Head Start REDI Program." *Child Development* 79(6): 1802-17.
- Berlinski, S, S. Galiani, and P. Gertler. 2009. "The Effect of Pre-primary Education on Primary School Performance." *Journal of Public Economics* 93 (1–2): 219–34.
- Berlinski, S., S. Galiani, and M. Manacorda. 2008. "Giving Children a Better Start: Preschool Attendance and School-Age Profiles." *Journal of Public Economics* 92 (5–6): 1416–40.
- Brown, J., S. Jones, M. LaRusso and L. Aber. 2010. "Improving Classroom Quality: Teacher Influences and Experimental Impacts of the 4Rs Program." *Journal of Educational Psychology* 102(1): 153-67.
- Chetty, R., J. N. Friedman, N. Hilger, E. Saez, D. W. Schanzenbach, and D. Yagan. 2011. "How Does your Kindergarten Classroom Affect your Earnings? Evidence from Project STAR." *Quarterly Journal of Economics* 126(4): 1593-1660.
- Chetty. R., J. N. Friedman, and J. Rockoff. 2013. "Measuring the Impact of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." NBER Working Paper 19423.
- Danielson, C. 1996. *Enhancing Professional Practice: A Framework for Teaching.* Alexandria, VA: Association for Supervision and Curriculum Development
- Domitrovich, C., S. Gest, S. Gill, K. Bierman, J. Welsh, and D. Jones. 2008. "Fostering High-Quality Teaching with an Enriched Curriculum and Professional Development Support: The Head Start REDI Program." *American Educational Research Journal* 46: 567-97.

References (continued)

- Downer, J., R. Pianta, X. Fan, B. Hamre, A. Mashburn, L. Justice. 2011. "Effects of Web-Mediated Teacher Professional Development on the Language and Literacy Skills of Children Enrolled in Prekindergarten Programs." *NSHA Dialog: A Research-to-Practice Journal for the Early Childhood Field* 14(4): 198-212.
- Downer, J., R. Pianta, M. Burchinal, S. Field, B. Hamre, J. LoSale-Crouch, C. Howes, K. LaParo and C. Scott-Little. 2013. "Coaching and Coursework Focused on Teacher-Child Interactions during Language-Literacy Instruction: Effects on Teacher Outcomes and Children's Classroom Engagement." Unpublished.
- Engle, P., M. Black, J. Behrman, M. Cabral de Mello, P. Gertler, L. Kapiriri, R. Martorell, and M. E. Young. 2007. "Strategies to Avoid the Loss of Developmental Potential in More Than 200 Million Children in the Developing World." *The Lancet* 369(9557): 229-42.
- Engle, P., L. Fernald, H. Alderman, J. Behrman, C. O'Gara, A. Yousafzai, M. Cabral de Mello, M. Hidrobo, N. Ulkuer, I. Ertem, S. Iltus, and the Global Child Development Steering Group. 2011. "Strategies for Reducing Inequalities and Improving Developmental Outcomes for Young Children in Low and Middle Income Countries." *The Lancet* 378(9799): 1339-53
- Gertler, P., J. Heckman, R. Pinto, A. Zanolini, C. Vermeersch, S. Walker, S. Chang, S. Grantham-McGregor. 2013. "Labor Market Returns to Early Childhood Stimulation: A 20-Year Follow-Up to an Experimental Evaluation in Jamaica." NBER Working Paper 19185.
- Grantham-McGregor, S., Y. B. Cheung, S. Cueto, P. Glewwe, L. Richter, B. Strupp and the International Child Development Steering Group. 2007. "Developmental Potential in the First 5 Years for Children in Developing Countries." *The Lancet* 369(9555): 60-70.
- Grantham-McGregor, S., C.. Powell, S. Walker, and J.. Himes. 1991. "Nutritional Supplementation, Psychosocial Stimulation, and Mental Development of Stunted Children: The Jamaican Study." *Lancet* 338 (8758): 1–5.
- Grantham-McGregor, S., S. Walker, S. Chang, and C. Powell. 1997. "Effects of Early Childhood Supplementation With and Without Stimulation on Later Development in Stunted Jamaican Children." *American Journal of Clinical Nutrition* 66 (2): 247–53.
- B. Hamre, R. Pianta, M. Burchinal, S. Field, J. LoSale-Crouch, J. Downer, C. Howes, K. LaParo and C. Scott-Little. 2012. "A Course on Effective Teacher-Child Interactions: Effects on Teacher Beliefs, Knowledge and Observed Practice." *American Educational Research Journal* 49(1): 88-123.
- Hanushek, E., and S. Rivkin. 2012. "The Distribution of Teacher Quality and Implications for Policy." *Annual Review of Economics* 4: 131-57.
- Imbens, G., and J. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62: 467-75.

References (continued)

- Jacob, B., and L. Lefgren. 2008. "Can Principals Identify Effective Teachers? Evidence on Subjective Performance Evaluation in Education." *Journal of Labor Economics* 26: 101-36.
- Kane, T., J. Rockoff, and D. Staiger. 2008. "What does Certification Tell Us About Teacher Effectiveness? Evidence from New York City." *Economics of Education Review* 27: 615-31.
- Kane, T., and D. Staiger. 2002. "Volatility in School Test Scores: Implications for Test-Based Accountability Systems." In D. Ravitch (ed.), Brookings Papers on Education Policy 2002, Brookings Institution, Washington, D.C.
- Kane, T., and D. Staiger, 2008. "Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation." NBER Working Paper 14607.
- Kane, T., and D. Staiger 2012. *Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains.* Bill & Melinda Gates Foundation.
- Kling, J., J Liebman, and L. Katz, 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica* 75(1): 83-119.
- Krueger, A. 1999. "Experimental Estimates of Education Production Functions." *Quarterly Journal of Economics* 114(2): 497-532
- Krueger, A., and D. Whitmore. 2001. "The Effect of Attending a Small Class in the Early Grades on College Test-Taking and Middle-School Test Results: Evidence from Project STAR." *Economic Journal* 11(468): 1-28.
- Macours, K., N. Schady, and R. Vakis. 2012. "Cash Transfers, Behavioral Changes, and Cognitive Development in Early Childhood: Evidence from a Randomized Experiment." *American Economic Journal: Applied Economics* 4(2): 247-73.
- Mashburn, A., R. Pianta, B., Hamre, J. Downer, O. Barbarin, D. Bryant, M. Burchinal, and D. Early. 2008. "Measures of Classroom Quality in Prekindergarten and Children's Development of Academic, Language, and Social Skills." *Child Development* 79(3): 732-49.
- Paxson, C., and N. Schady. 2007. "Cognitive Development among Young Children in Ecuador: The Roles of Wealth, Health, and Parenting." *Journal of Human Resources* 42(1): 49-84.
- Paxson, C., and N. Schady. 2010. "Does Money Matter? The Effects of Cash Transfers on Child Development in Rural Ecuador." *Economic Development and Cultural Change* 59(1): 187-229.
- Pianta, R.C. 2011. "Teaching Children Well: New Evidence-Based Approaches to Teacher Professional Development and Training." Center for American Progress.
- R. Pianta, A. Mashburn, J. Downer, B. Hamre, L. Justice. 2008. "Effects of web-mediated professional development resources on teacher-child interactions in pre-kindergarten classrooms." *Early Childhood Research Quarterly* 23: 431-51.

References (continued)

- Pianta, R.C., and B. Hamre. 2009. "Conceptualization, Measurement, and Improvement of Classroom Processes: Standardized Observation can Leverage Capacity." *Educational Researcher 38*(2): 109-119.
- Pop-Eleches, C., and M. Urquiola. 2013. "Going to a Better School: Effects and Behavioral Responses." Forthcoming, *American Economic Review*.
- Rivkin, S., E. Hanushek, and J. Kain. 2005. "Teachers, Schools, and Academic Achievement." *Econometrica* 73: 417-58.
- Rockoff, J., and D. Staiger. 2010. "Searching for Effective Teachers with Imperfect Information." *Journal of Economic Perspectives* 24(3): 97-118.
- Romano, J. P., and M. Wolf. 2005. "Stepwise Multiple Testing as Formalized Data Snooping." *Econometrica* 73: 1237–1282.
- Rothstein, J. 2010. "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement." *Quarterly Journal of Economics* 125(1): 175-214.
- Schady, N. 2011. "Parental Education, Vocabulary, and Cognitive Development in Early Childhood: Longitudinal Evidence from Ecuador." *American Journal of Public Health* 101(12): 2299-307.
- Schady, N., J. Behrman, C. Araujo, R. Azuero, R. Bernal, D. Bravo, F. Lopez-Boo, K. Macours, D. Marshall, C. Paxson, and R. Vakis. 2012. "Wealth Gradients in Early Childhood Cognitive Development in Five Latin American Countries." Paper presented at "Promises for Preschoolers: Early Childhood Development and Human Capital Accumulation" Conference, University College London.
- Staiger, D., and J. Rockoff. 2010. "Searching for Effective Teachers with Imperfect Information." *Journal of Economic Perspectives* 24(3): 97-118.
- Walker, S., S. Chang, C. Powell, and S. Grantham-McGregor. 2005. "Effects of Early Childhood Psychosocial Stimulation and Nutritional Supplementation on Cognition and Education in Growth-Stunted Jamaican Children: Prospective Cohort Study." *Lancet* 366 (9499): 1804–07.
- Walker, S., S. Grantham-McGregor, C. Powell, and S. Chang. 2000. "Effects of Growth Restriction in Early Childhood on Growth, IQ, and Cognition at Age 11 to 12 Years and the Benefits of Nutritional Supplementation and Psychosocial Stimulation." *Journal of Pediatrics* 137(1): 36–41.
- Yoshikawa, H., C. Snow, C. Weiland, D. Leyva, M.C. Arbour, C. Barata, E. Treviño, S. Mendive, and A. Rolla. 2013. "Improving the Quality of Early Childhood Education in a Middle-Income Nation: Evidence from Chile." Unpublished manuscript, Harvard University.

Battery of tests

- 12 tests (TVIP, Woodcock-Johnson battery, Stroop test, multi-dimensional card sort)
- Piloted and validated in Ecuador
- Approximately 40 minutes per child

3 areas:

- **Language and early literacy**
    1. Letter & word id
    2. Id of first sound of words
    3. Receptive vocabulary (TVIP)
    4. Oral comprehension
- **Math**
    1. Number ID
    2. Block rotation
    3. Numeric series
    4. Applied problems
- **Executive function**
    1. Inhibitory control
    2. Attention control
    3. Cognitive flexibility
    4. Working memory

Language and early literacy

1. Letter word ID:

la **pan** **de** **tren**

*Señala "**DE**"*

Woodcock-Muñoz, Batería III (Woodcock, McGrew & Schenk, 2007)

Language and early literacy

2. Recognition of the sound of letters and words:

¿Cuál es el sonido de esta letra?

F

M

a

P

¿Y de esta otra?

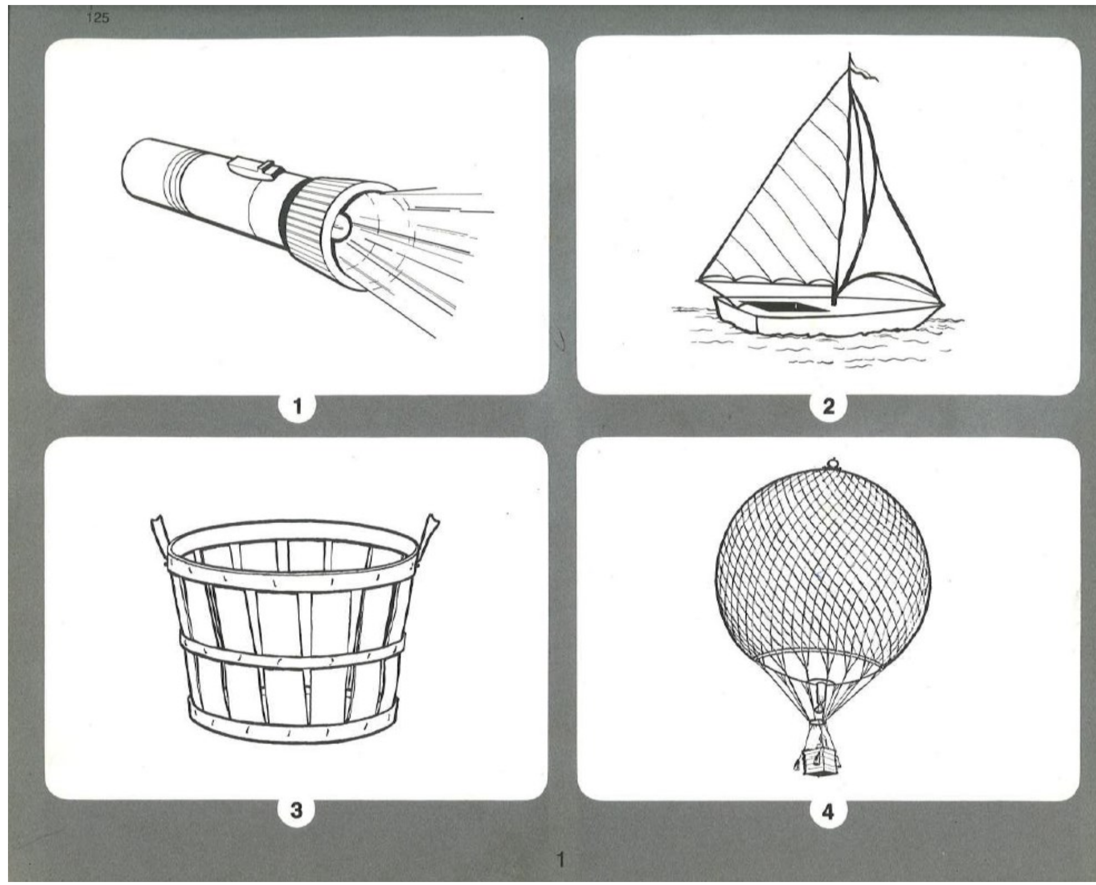
Language and early literacy

2. Recognition of the sound of letters and words:

Language and early literacy

3. Receptive language (TVIP)

Señala
BARCO



Language and early literacy

4. Oral Comprehension

Pablo mira la televisión todos los días. Su programa favorito es el de dibujos animados.

¿Cómo se llama el niño?

Math

1. Number recognition

¿Qué número es este?

4

2

7

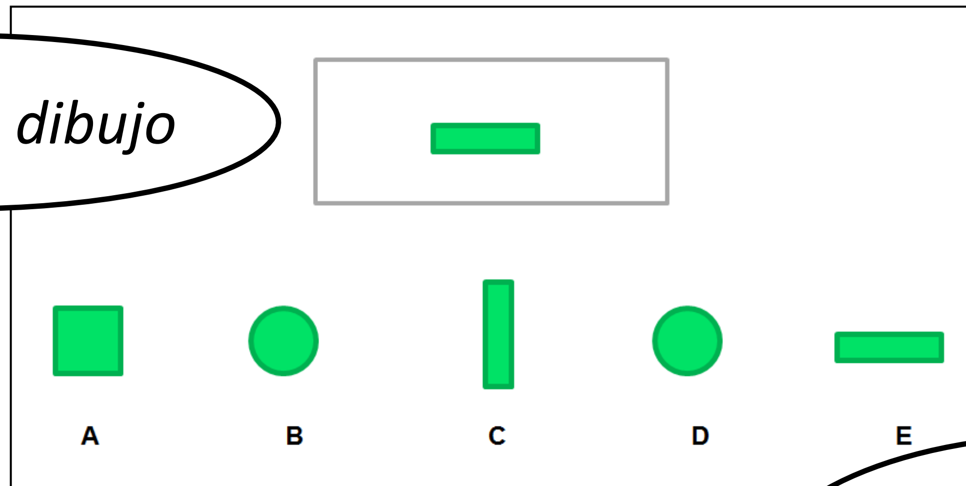
12

¿Y es este otra?

Math

2. Block rotation

Mira este dibujo



¿Cuáles dos de estos dibujos son iguales a este?

Woodcock-Muñoz, Batería III (Woodcock, McGrew & Schenk, 2007)

Math

3. Numeric series

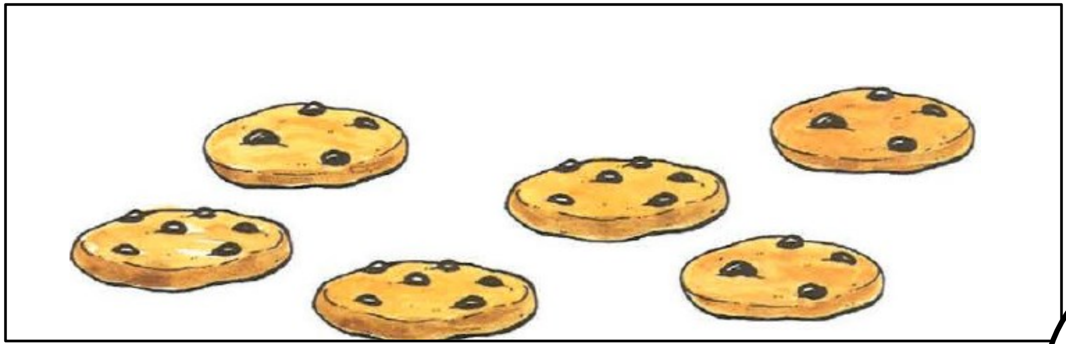
Siete, un número que falta, nueve y diez

7		9	10

¿Cuál es el número que falta?

Math

4. Applied problems



Si José se come tres galletas, ¿cuántas galletas quedan?

Executive function

1. **Inhibitory control**: (Stroop Day & Night)

*Cuando veas este dibujo quiero que digas "**DÍA**"*

Executive function

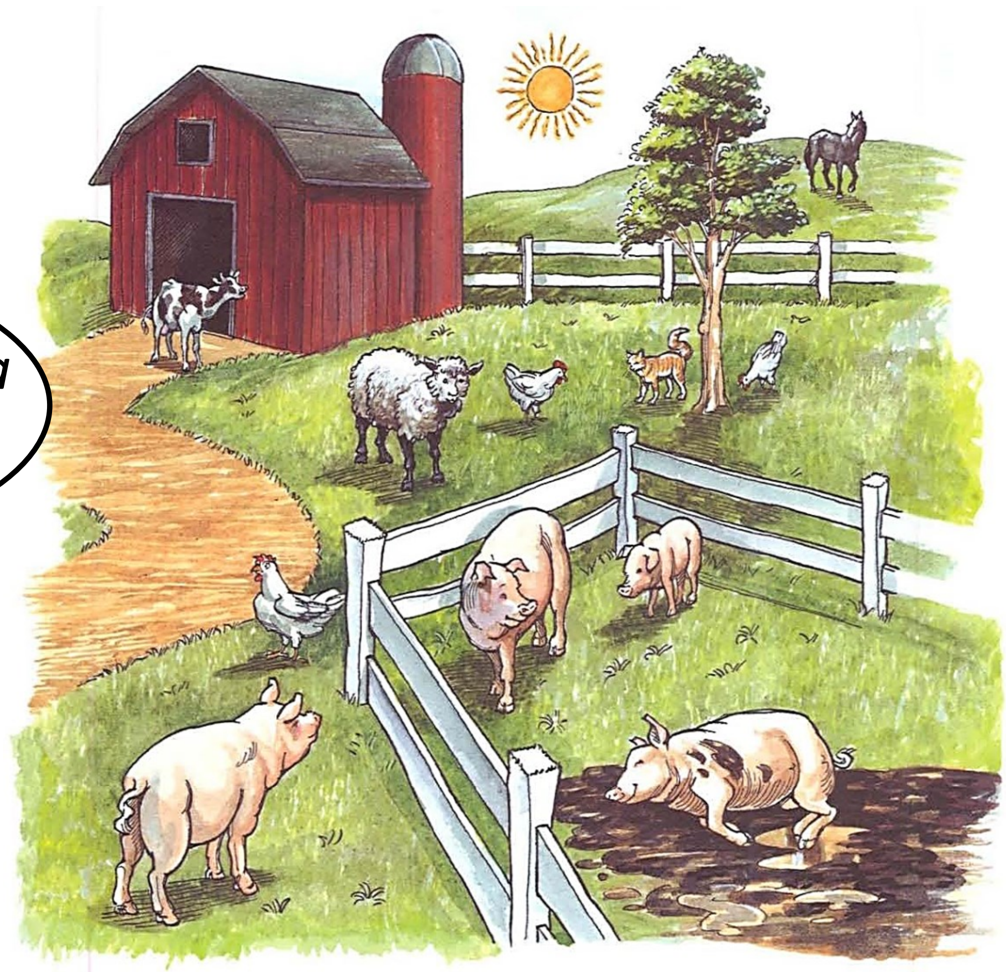
1. Inhibitory control: (Stroop Day & Night)

*Cuando veas este dibujo quiero que digas "**NOCHE**"*

Executive function

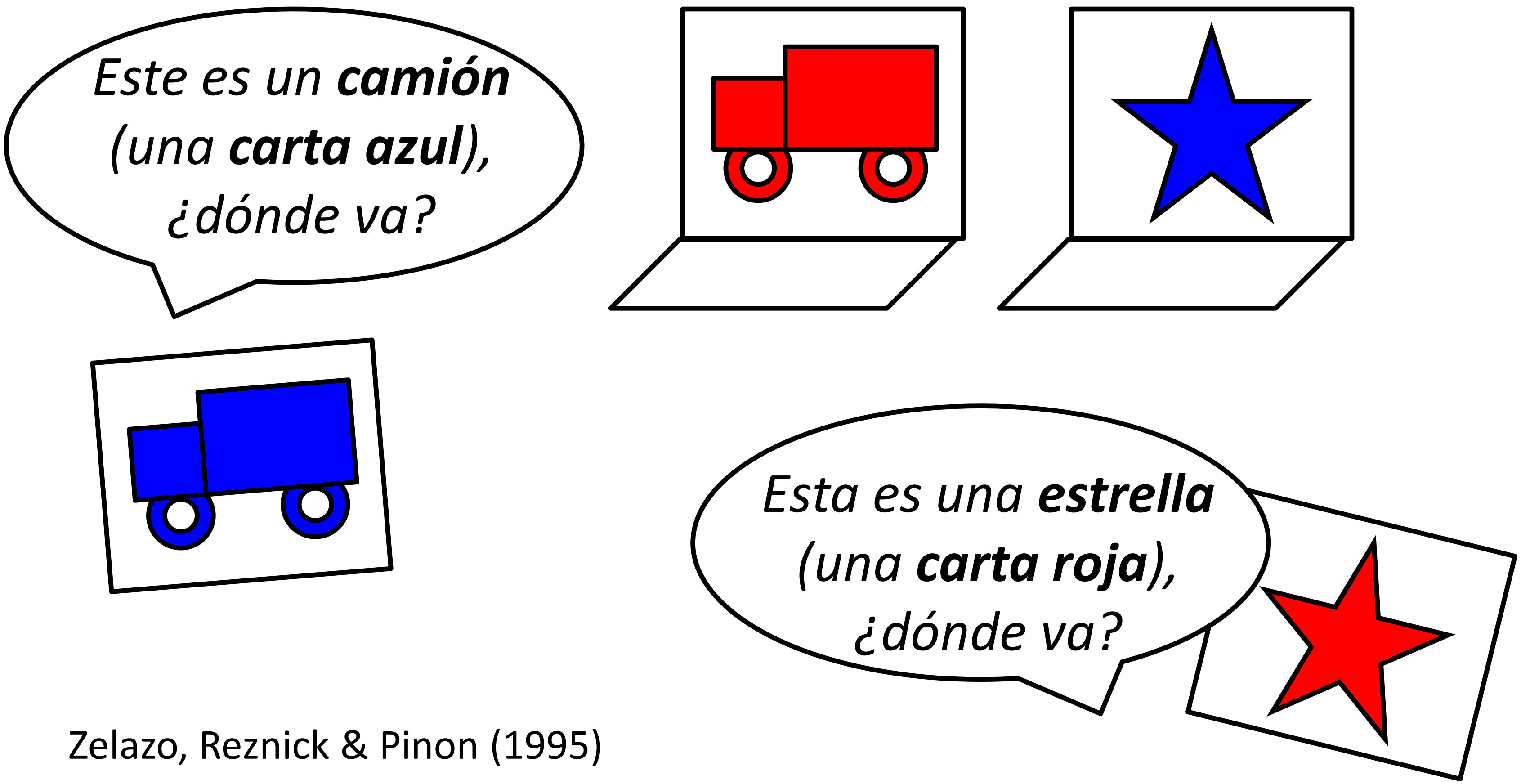
2. Attention control

Señala el árbol, luego la vaca y por último el sol



Executive function

3. Cognitive flexibility: (Multidimensional card sort)



Zelazo, Reznick & Pinon (1995)

Executive function

4. Working memory:

Te voy a decir unos números y quiero que los repitas en el mismo orden

Tres, seis, cinco, uno

Repite

Did the experiment “work”?

	CLASS, 2011/12	Experience (in years)
Nominally enrolled, never attended *N* = 241	0.0012 (0.0022)	-0.0002 (0.0003)
Not assigned (late enrollment) *N* = 261	-0.0002 (0.0025)	0.0000 (0.0002)
Missing baseline TVIP *N* = 606	-0.0053 (0.004)	0.0002 (0.0004)
Missing other covariate information *N* = 1,134	0.0035 (0.0041)	-0.0002 (0.0006)
Child dropped out *N* = 187	-0.0012 (0.0032)	-0.0004 (0.0003)
Missing end-of-year test data *N* = 352	0.0051 (0.0033)	0.0003 (0.0005)

Notes: Table reports the coefficient from a separate OLS estimation of each explanatory variable (second and third columns) on the dependent variable, given in each row of the table. All regressions include school fixed effects. Standard errors in parentheses, clustered at school level.
** significant at 5% level, *** at 1% level.