A Helping Hand? Teacher Quality and Learning Outcomes in Kindergarten

Maria Caridad Araujo

Pedro Carneiro

Yyannú Cruz-Aguayo

Norbert Schady

Motivation

- We seek to answer a simple, but critical question: How much, and in what ways, do kindergarten teachers matter for learning outcomes?

- Huge literature on teachers from developed and developing countries (beginning with Hanushek 1971 and Murnane 1975):

  - <u>United States</u>: important contributions include Chetty et al. (2011), Chetty et al. (2014), Hanushek and Rivkin (2012), Jacob and Lefgren (2008), Kane and Staiger (2008), Krueger (1999), McCaffrey et al. (2009), Nye et al. (2009), Rivkin et al. (2006), Rothstein (2010), Staiger and Rockoff (2010), among others

  - <u>Developing countries</u>: important contributions include Andrabi et al. (2011), Duflo et al. (2011, 2012), Glewwe et al. (2011, 2012), Kremer and Holla (2009), Kremer et al. (2013), among others

Motivation

Two challenges:

1. Convincingly evaluating the causal effect of teachers on child learning is difficult because of sorting, possibly on unobservables (which may be time-varying)
   - Solution:
     - Randomly assign **two** cohorts, totaling ~23,000 children entering kindergarten, to teachers within 204 schools in the coastal region of Ecuador
     - Compliance with random assignment>98 percent
     - Also collect data on nationally representative sample of 100 schools (to check for external validity)

2. Data that are generally available on teachers (degrees, experience, tenure status) are only very weakly correlated with performance
   - Solution: collect very rich data on teachers:
     - Teacher characteristics: experience, tenure status, IQ, Big Five, Attention & Inhibitory Control, early circumstances
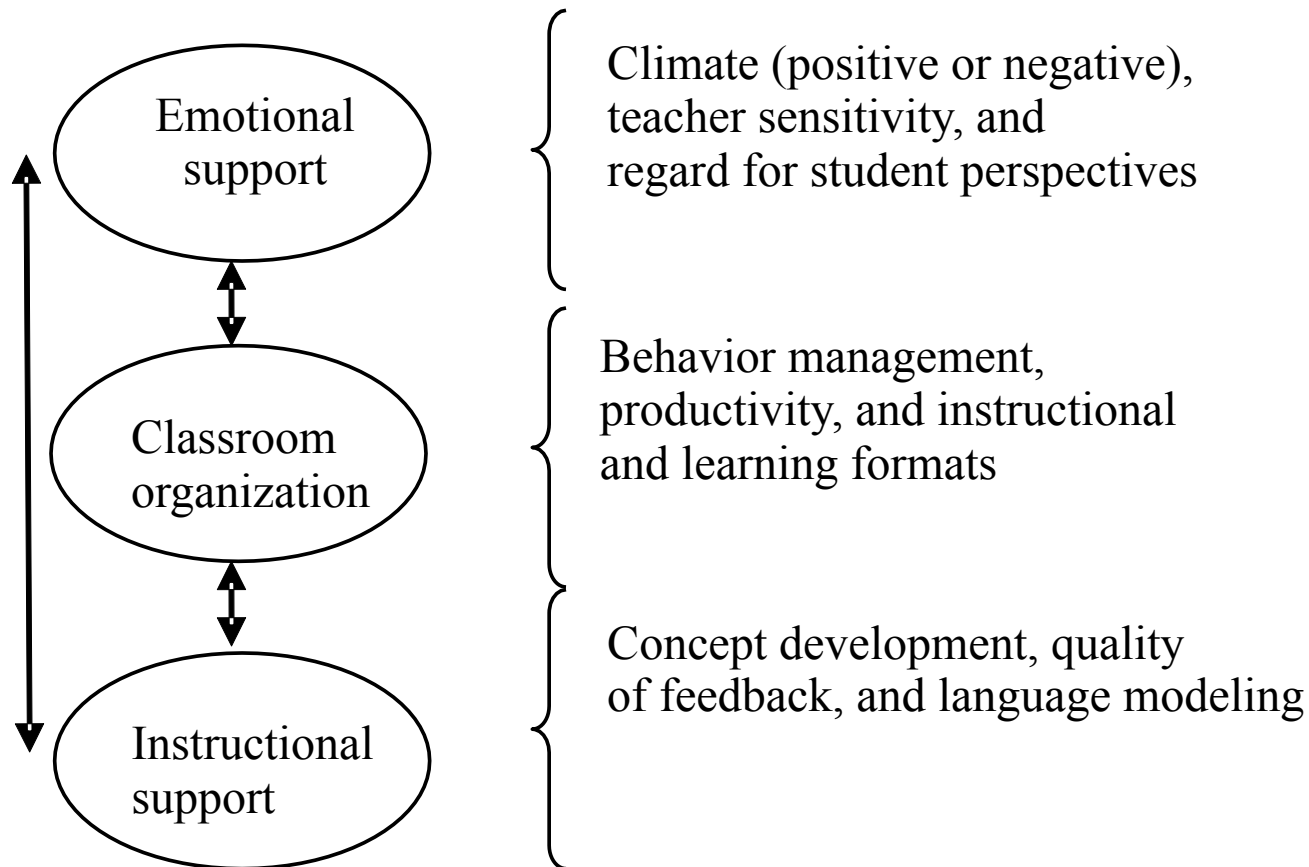   - Teacher behaviors: CLASS, absenteeism

Study design

- Collect very rich data on children:
  - Child receptive vocabulary, age, gender at baseline
  - Test <u>math</u> (4 tests), <u>language</u> (4 tests) and <u>executive function</u> (EF) (4 tests) at follow-up
  - 92.5 percent of all children completed all tests
    - Test completion is not associated with teacher characteristics: no obvious concerns about selection
  - Executive function is a set of basic self-regulatory skills which involve various parts of the brain, but in particular the prefrontal cortex
    - *Inhibitory control* refers to the ability to suppress impulsive behaviors and resist temptations
    - *Working memory* refers to the ability to hold, update, and manipulate information in the mind for short periods of time
    - *Cognitive flexibility* refers to the ability to shift attention between competing tasks or rules
    - *Attention* is the ability to focus and disregard external stimuli
  - EF is an important determinant of how well young children adapt to and learn in school, and succeed in life (Moffitt et al. 2011)

Study design

- Collect very rich data on teachers:
  - Data on teachers that are generally available in administrative data: age, gender, years of experience, whether tenured or working on a contract basis
  - Additional data on teacher characteristics: teacher IQ, personality (as measured by the Big Five), teacher executive function, early circumstances
  - Data on teacher behaviors: teacher absenteeism, and teaching practices, as measured by the CLASS (the Classroom Assessment Scoring System) (Pianta et al. 2007)
    - For the purpose of the CLASS, all teachers in the study were filmed for an entire school day, and the resulting video was double-coded by coders expressly trained for this purpose (>4900 segments)
  - Some attrition of teachers (because they change schools): 53 teachers left the schools in our study sample *between* the 2011/12 and 2012/13 school years, and 72 moved *within* the 2012/13 year
    - However, teachers were no more likely to move within the school year if, by chance, they were assigned a particularly difficult group of students

What is the CLASS and why use it?

Classroom observation tool

Emotional support

Climate (positive or negative), teacher sensitivity, and regard for student perspectives

Classroom organization

Behavior management, productivity, and instructional and learning formats

Instructional support

Concept development, quality of feedback, and language modeling

Example: Teacher Behaviors and CLASS Scores for Behavior Management Dimension

	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.

Source: Pianta, La Paro & Hamre (2008)

CLASS scores
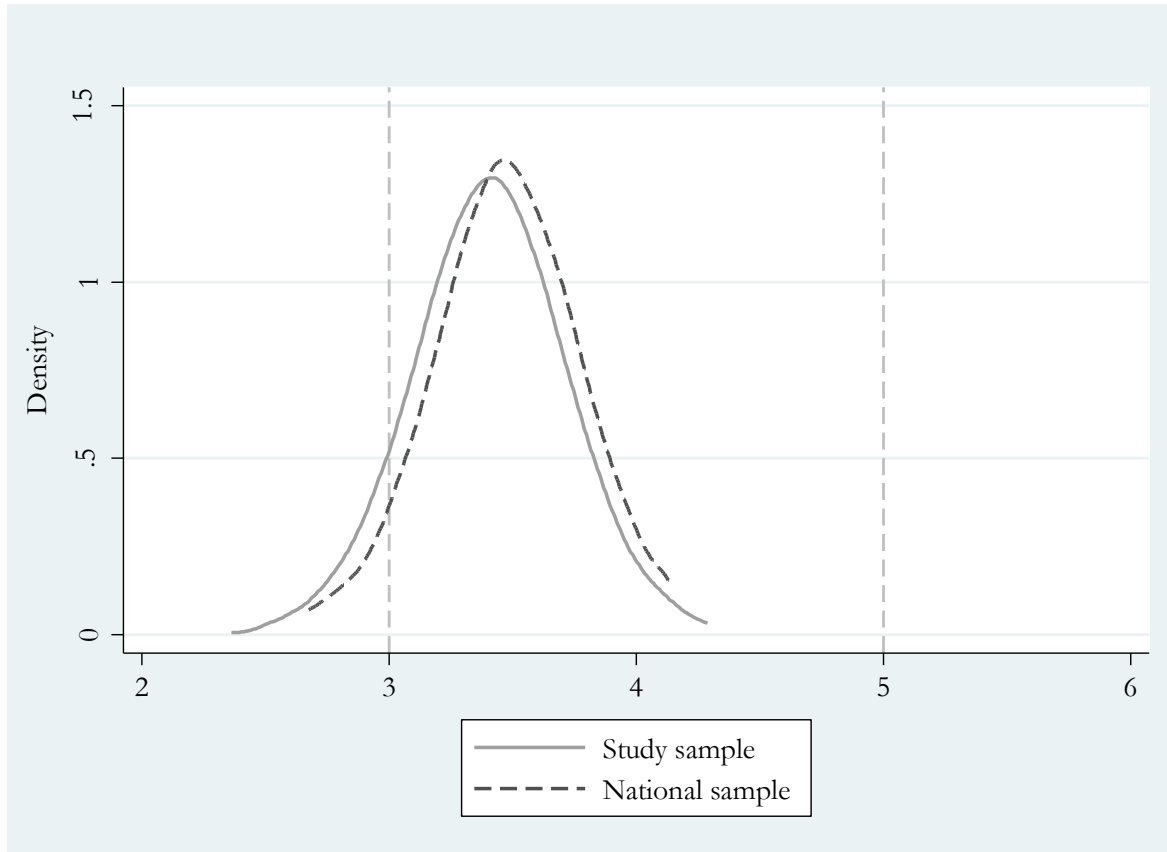
~50 percent of variation in CLASS scores across teachers within schools, ~50 percent across schools

Estimation strategy

- Two parts to analysis:
  1. Estimate teacher effects: How much does it matter whether a child was assigned to teacher A or B in a school? This analysis calculates the <u>total</u> teacher effect, but says nothing about *why* or *how* one teacher is better than another
     - How to scale this? Set teacher effect <u>within</u> each school equal to zero, calculate standard deviation of demeaned teacher effects
     - Correct for measurement error using techniques that are standard in the literature: Empirical Bayes , covariance in teacher effects across years (Kane and Staiger 2002; Chetty et al. 2011; Hanushek and Rivkin 2012)
  2. Estimate the associations between within-school differences in teacher characteristics or behaviors and child learning outcomes
     - May or may not be causal
     - Great deal of measurement error in CLASS: Instrument CLASS in year t with CLASS for same teacher in t-1 (with *different* cohort of students)

Results: Teacher Effects

1. After correcting for sampling error, a one-sd increase in teacher quality leads to increases in child learning of 0.11 sds in math, 0.13 sds in language and 0.07 sds in executive function

   – Results for language and math very close to those reported in the US literature (summaries in Hanushek and Rivkin 2012, Kane and Staiger 2010)

		2012/13 (12 tests)		2012/13 (4 tests)		2013/14 (4 tests)		Covariance, 2012/13 & 2013/14
		U	C	U	C	U	C	
Language		0.17	**0.13**	0.14	0.10	0.15	0.11	0.09
Math		0.16	**0.11**	0.16	0.11	0.16	0.11	0.09
EF		0.14	**0.07**	--	--	--	--	
TOTAL		0.15	**0.11**	0.16	0.11	0.15	0.11	0.10
Schools		150		87		87		87
Teachers		334		196		196		196
Children		10,465		6,213		6,170		

Table 1: Teacher effects

Note: The values reported are the standard deviation of the classroom averages. In the specification that includes all 12 tests, controls include all baseline child and household characteristics. All other specifications controls include only child age, gender, and the baseline TVIP score.
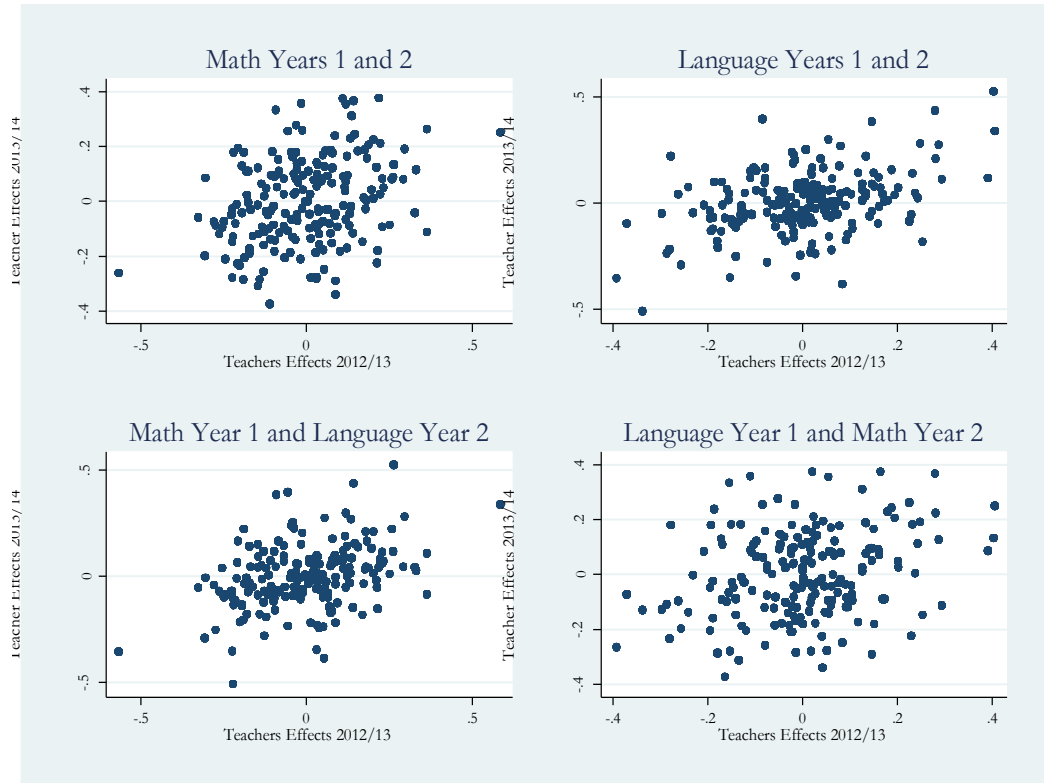
Results: Teacher Effects

How big are these teacher effects?

- Being assigned to an outstanding (a teacher at the 95th percentile of the distribution in a school), rather than an average teacher (a teacher at the 50th percentile of the distribution) has an effect on test scores that is equivalent to ~1/4 of the test score gap between children of mothers who are secondary school graduates and those who are elementary school dropouts

- Being assigned to an outstanding, rather than an average teacher, improves test scores by ~magnitude as 18 months of cash transfers, equivalent to 10 percent of mean household per capita expenditures (0.17 sds, as reported by Paxson and Schady 2009 for Ecuador) or to the effect of attending kindergarten (0.23 sds, as reported by Berlinski et al. 2009 for Argentina)

  – Because we ignore cross-school differences in teacher quality, are probably under-estimating total variation in teacher quality by ~ a factor of two

Results: Teacher Effects

2. The *same* teachers have their children learn more math *and* more language year after year

- Cross-year correlation of teacher effects in math: 0.32 (similar to estimates from LA and NYC, see Kane and Staiger 2010)

- Cross-year correlation of teacher effects in language: 0.42 (similar to estimates from LA and NYC, see Kane and Staiger 2010)

- **Cross-year, cross-subject correlations** of teacher effects: 0.28, 0.37

Results: Teacher characteristics and behaviors

- In Ecuador, as elsewhere, there are returns to experience which rise sharply in the first two years and then flatten out: Children randomly assigned to "rookie" teachers learn less (0.16 sds)
- Children randomly assigned to teachers with higher IQ learn more, although the effect is modest (0.04 sds)
- Children randomly assigned to teachers with higher CLASS scores learn <u>substantially</u> more:
  - In OLS regressions, without correcting for measurement error, a one-point increase in the CLASS is associated with 0.19 sds more learning
  - In IV regressions, which correct for measurement error, a one-point increase in the CLASS is associated with 0.59 sds more learning
- Teachers with better CLASS scores get **all** their students to learn more (not shown): Effects are not concentrated on girls or boys, on children with high or low levels of development when they enter school, or on children of high or low socioeconomic status
  - "A rising tide lifts all boats"

Results: Teacher characteristics

Table 2: Teacher effects, characteristics available in administrative data						
	2012/13 cohort					2013/14 cohort
	Language	**Math**	**Executive function**	**All**	**Common tests**	**Common tests**
Inexperienced ("rookie") teachers	-.130 (.086)	-.153 (.089)	**-.115 (.058)**	**-.159 (.077)**	-.157 (.089)	
Tenured	.051 (.041)	.101 (.043)	.013 (.041)	.062 (.044)	.086 (.044)	.075 (.039)

Note: All regressions include baseline student and household characteristics, their classroom averages, and school fixed effects. Standard errors clustered at the school level. Coefficients that are significant at the 5 percent level are in bold.

Results: Teacher characteristics

		2012/13 cohort					2013/14 cohort
Table 3: Teacher effects, additional characteristics							
		Language	**Math**	**Executive function**	**All**	**Common tests**	**Common tests**
IQ		**.043 (.021)**	**.038 (.018)**	.023 (.016)	**.041 (.018)**	.032 (.018)	.016 (.018)
Big Five							
	Neuroticism	-.003 (.017)	-.003 (.017)	.011 (.016)	.001 (.016)	-.006 (.016)	.007 (.021)
	Extraversion	.026 (.019)	.031 (.019)	.024 (.013)	.032 (.016)	.032 (.017)	.003 (.015)
	Openness	.007 (.021)	.018 (.021)	.008 (.017)	.013 (.020)	.022 (.021)	.009 (.018)
	Agreeableness	-.001 (.022)	.000 (.022)	-.016 (.017)	-.006 (.021)	-.009 (.020)	.002 (.021)
	Conscientiousness	-.025 (.018)	-.035 (.019)	-.023 (.016)	-.032 (.018)	-.033 (.019)	-.030 (.020)
Inhibitory control		.024 (.029)	.034 (.017)	**.032 (.014)**	.035 (.020)	.024 (.018)	.008 (.022)
Parents' schooling		.006 (.006)	.007 (.005)	.003 (.005)	.006 (.005)	.004 (.005)	-.004 (.005)

Note: All regressions include baseline student and household characteristics, their classroom averages, and school fixed effects. Standard errors clustered at the school level. Coefficients that are significant at the 5 percent level are in bold.

Results: Teacher behaviors

	2012/13 cohort					2013/14 cohort
Table 4: Teacher effects, behaviors						
	Language	**Math**	**Executive function**	**All**	**Common tests**	**Common tests**
Teacher attendance (>3 days of absence)	.028 (.050)	-.042 (.034)	-.059 (.039)	-.027 (.037)	-.013 (.034)	**-.089 (.040)**
Lagged CLASS	**.182 (.082)**	**.192 (.078)**	**.121 (.060)**	**.194 (.075)**	**.184 (.081)**	**.136 (.045)**

Note: All regressions include baseline student and household characteristics, their classroom averages, and school fixed effects. Standard errors clustered at the school level. Coefficients that are significant at the 5 percent level are in bold.

Results: Teacher behaviors

Table 5: CLASS effects, with and without correction for measurement error									
		Language		Math		Executive function		All	
		OLS	IV	OLS	IV	OLS	IV	OLS	IV
Lagged CLASS		**.221 (. 064)**	**.554 (. 238)**	**.232 (. 065)**	**.589 (. 237)**	.078 (. 054)	**.368 (.180)**	**.211 (.060)**	**.593 (.225)**

Note: All regressions include baseline student and household characteristics, their classroom averages, and school fixed effects. Standard errors clustered at the school level. Coefficients that are significant at the 5 percent level are in bold.

Additional results (in paper)

1. Parents have a good sense of who are better and worse teachers. On a five-point scale:
   - Teachers who produce 1 sd more learning for class are given a 0.44 point higher score
   - Rookie teachers are given a 0.33 point lower score
   - Teachers with a 1-point higher (lagged) CLASS score are given a 0.12 point higher score

2. However, parents do **not** adjust their behaviors in response to differences in teacher quality:
   - There is no effect of teacher quality on dropout or child absenteeism: Teacher effects are working on the <u>intensive margin</u> (more learning per day of school) rather than the <u>extensive margin</u> (more days of school)
   - Parents do not change the amount they read to, play with, tell stories to or any one of a number of behaviors in response to differences in teacher quality: They do not engage in <u>compensatory</u> or <u>reinforcing</u> behavior (unlike the results in Pop-Eleches and Urquiola 2013 for Romania)

Policy simulations

1. Teacher deselection: Replace the worst-performing 5 percent (10 percent) of teachers in year *t* with a teacher whose value added is the average for her school: this would increase overall learning outcomes by 0.07 (0.11) standard deviations in year *t+1*

2. Teacher re-training: Choose the 20 percent (40 percent) of teachers with the lowest CLASS scores and have these teachers participate in a program that raises their CLASS scores by 0.5 points (reasonable from US literature): this policy would increase overall child learning outcomes by 0.06 (0.13) standard deviations in year *t+1*
    – Caveat that CLASS "effects" may not be causal

Conclusions

- In middle income countries, **access** to schooling for children of elementary school age is no longer a major challenge
- However, the **quality** of education is frequently dismal
  - Findings from various evaluations that have sought to improve quality by increasing inputs have been disappointing: Recent reviews:
    - Kremer et al. (2013) concludes that "providing additional inputs (to schools) without changing pedagogy has had little impact"
    - Murnane and Gaminian (2014) concludes that "more or better resources do not improve student achievement unless they change children's daily experiences at school"
- We use a convincing experimental design to show that:
  1. Even within schools, there are substantial differences in teacher quality
  2. The observable characteristics of teachers explain very little of these differences
  3. However, classroom *practices* are important predictors of child learning

Extensions

- Kindergarten children re-randomized to first grade, second grade, third grade teachers
    - Can estimate fade-out, and experimental estimates of dynamic complementarities
- Pilot of in-service training and mentoring about to go into the field
- Lots of other work underway
    - Peer effects (and how they interact with the CLASS)
    - Can kindergarten children recognize better teachers?

Thank you