Levers for Learning:
Relationships between school-level factors and literacy outcomes in low-income schools in Colombia[1]

Felipe Barrera-Osorio

Andreas de Barros

Sarah Dryden-Peterson

Bethany Mulimbi

Nozomi Nakajima

Paola Uccelli

Citation:
Barrera-Osorio, F., de Barros, A., Dryden-Peterson, S., Mulimbi, B., Nakajima, N., & Uccelli, P. (2020). Levers for Learning: Relationships between School-Level Factors and Literacy Outcomes in Low-Income Schools in Colombia. *Comparative Education Review*, *64*(2).

Abstract:
This study examines the relationships between school processes and literacy in public schools that serve low-income families in a mid-size city in Colombia. We focus on four categories of school processes: instructional practice; school-community engagement; well-being of students, teachers, and parents; and community belonging. We present two key results. First, we find considerable variation in literacy performance across schools, despite serving students with similar family background characteristics. Second, we find community belonging to be the most salient school process for predicting literacy. Among measures of community belonging, higher levels of students' sense of belonging and lower levels of bullying were associated with higher literacy. This study contributes to the field of international and comparative education by highlighting the importance of relational and interpersonal dimensions of school processes for children's learning. Our study also demonstrates the value of adopting an interdisciplinary lens, combining economics, child development, and sociology, to better understand the links between school processes and student learning.

Keywords:
education, low- and middle-income countries, literacy, differentiated instruction, community engagement, well-being, belonging, bullying

[1] This project would not have been possible without the help and support of the Luker Foundation. We are extremely grateful for the support and work of Santiago Isaza and Camila Arango. We also thank the schools in the city of Manizales. This Project was conducted with financial support from the Dean's Venture Fund, Harvard Graduate School of Education.

Introduction

School enrollment rates have increased rapidly in low- and middle-income countries over recent years. This success is accompanied by a serious challenge: schools continue to struggle with how to serve all students to achieve meaningful learning, including the skills, knowledge, values, and attitudes to prepare them to be engaged members of society who participate actively economically and civically (Pritchett, 2013; World Bank, 2017; OECD, 2016; Raitano & Vona, 2013; Wößmann, 2005).

This research study is part of a larger project that seeks to identify relatively "high-performing schools" among public schools that serve low-income communities in low- and middle-income countries, and to investigate the characteristics that distinguish them from relatively "low-performing schools" that are otherwise comparable, including the populations they serve. Through this comparison, we attempt to identify what we call "levers for learning," which are school-level processes that predict school success.[2] Given rapid urbanization globally, this project examines public schools in urban settings (UNICEF, 2012). We also focus on students' literacy as a critical foundational skill for academic achievement and engaged citizenship. In the present study, we collected data in ten schools – five relatively "high-performing" and five relatively "low-performing" institutions – in the city of Manizales, Colombia. These data were collected building on partnerships with a local NGO and with the Secretary of Education, the authority in charge of public education in the city.

In the comparative education literature, the contributions of system-level policies, school input measures, and individual-level factors to school learning are well researched; yet, interpersonal and relational dimensions of school processes (i.e., dimensions related to

[2] We purposefully use language that echoes "points of leverage" that Schiefelbein and Farrell invoked at a systems-level in their ground-breaking longitudinal study of students in Chile (Schiefelbein and Farrell, 1982, p. 24).

teachers-to-students, teachers-to-parents, students-to-students interactions) have been less studied. Research on the types of system-level policies that lead to better educational outcomes is extensive and illustrated, for instance, by the multiple cross-country studies based on the PISA assessment data (e.g., OECD, 2012, 2016). Another approach, followed by earlier studies in low- and middle-income countries, used convenience datasets to explore both school input measures and students'/households' characteristics that are associated with student achievement (Heyneman and Loxley, 1982, 1983a, 1983b; more recent estimations include Fuller, 1987; Baker, Goesling and LeTendre, 2002; Chiu & Khoo, 2005). In contrast to these approaches, our study explores relational and interpersonal dimensions of *school processes*. In other words, we focus on measures that capture what goes on regularly in school interactions that involve students, teachers, and parents. We hypothesize that these school processes explain some of the variation in student achievement across schools that serve similar populations from low-income neighborhoods. Specifically, we investigate four levers for learning: instructional practices; school-community engagement; well-being of students, teachers, and parents; and community belonging in a typical mid-size city in Colombia. With this aim, we collected data to serve as proxy of these levers and to explore each of them in relation to student literacy achievement.[3]

Colombia provides a productive site for our inquiry as it represents a common paradox in low- and middle-income countries: students in public schools exhibit low learning outcomes despite important investments in public education. For instance, even though 15 percent of the national expenditure went to education in 2015 (World Bank, 2015, 2017), the average score of Colombia in the PISA 2015 literacy test was 425 points, 68 points below the OECD average (493). Moreover, 43 percent of Colombian students scored at PISA level 1 or below the baseline level, which indicates minimal or no reading comprehension (OECD, 2016). Our data come from

[3] Closer to this approach are Carnoy & Marshall (2005) and Carnoy et al. (2015)

Manizales, a typical Colombian mid-size city, with challenges related to student literacy learning.[4]

We conducted our analysis in two phases. First, we validated the research design. We tested whether the two types of schools – high-performing and low-performing – which we identified from existing achievement data, in fact (1) showed large differences in literacy scores based on the tests we administered, and, (2) served low-income populations with similar observable characteristics. Second, we used a basic statistical model to explore correlations between levers for learning and student literacy outcomes, after controlling for baseline household and student characteristics. It is important to state, though, that the nature of the data does not allow us to estimate causal models. We cannot rule out that non-observable characteristics of students may be different across schools. Instead, this is an exploratory study aimed to find schools serving similar students on observable characteristics. We examine correlations between levers and literacy outcomes, with the goal of shedding light on which school-level processes are associated with students' literacy.

To validate our research design, we first show that the schools in our study indeed have significant differences in students' literacy achievement despite serving demographically similar populations. This is in itself an important finding as it suggests that schools may contribute to learning outcomes for marginalized students. Through our second set of analyses, we found that the most salient lever for student literacy outcomes was community belonging. Specifically, two aspects of community belonging emerged as significant predictors of literacy: sense of school belonging and bullying. Moreover, results revealed that students in high-performing schools

[4] As any study in a specific location, extrapolation to the rest of the country is difficult to assess. Nevertheless, this study offers relevant inisghts that might inspire further reserach in other other mid-sized cities in the country and in other middle-income countries.

were more likely to exhibit a positive sense of school belonging and that high-performing schools also displayed significantly lower levels of bullying.

Conceptual Framework: Levers for Learning

In an education production function framework (Hanushek 1979, 2003), educational achievement is modeled as a function of individual, household, and community characteristics, as well as school inputs.[5] An early application of this framework was the Coleman et al. (1966) study, which demonstrated that individual and family characteristics explained a larger proportion of the variation in students' test scores relative to school characteristics in the United States. Heyneman & Loxley (1982, 1983a, 1983b), however, showed that in the context of low- and middle-income countries the variation in student achievement explained by school characteristics was larger than family socioeconomic status. Later research has demonstrated that this relative effect of school resources over family background on achievement holds across a range of low- and middle- income countries (Fuller, 1987; Baker, Goesling and LeTendre, 2002; Chiu & Khoo, 2005).

Many of these comparative education studies measure school characteristics using cross-national quantitative data. These characteristics are captured by school input measures, such as a principal's years of experience, the percent of teachers with five or more years of experience; or student input measures, such as the percent of students absent (Hanushek, 2003). Rather than focus on school or student input measures, our paper aims to explore school *processes*, using unique survey data collected in schools.[6] In particular, we focus on

[5] Todd & Wolpin (2003) present a formal analysis of the estimation of production functions in education.

[6] Toots & Lauri (2015) and Põder, Lauri & Veski (2017) are examples of studies that collect rich data at the school level and then aim to explore the relationship between a lever and student achievement, in the first case,

interpersonal and relational dimensions of what goes on inside schools regularly. Based on the literature reviewed below, we hypothesize that these school processes may promote literacy achievement and we derive four levers of learning.

In Figure 1, we present a conceptual model that guides our analysis. Given our design, we examined schools that share observable characteristics in terms of system-level policies and input resources (Box A). We also selected schools that serve populations of students who share similar observable characteristics (Box B). In our research, we examined differences among schools on four selected processes that take place within schools, what we call levers for learning (Box C). We examine these levers as hypothesized moderators of student achievement (Box D) in public schools serving similar low-income populations. In the next section, we explain the rationale for these levers.

---

civic knowledge, in the second, selection of schools.

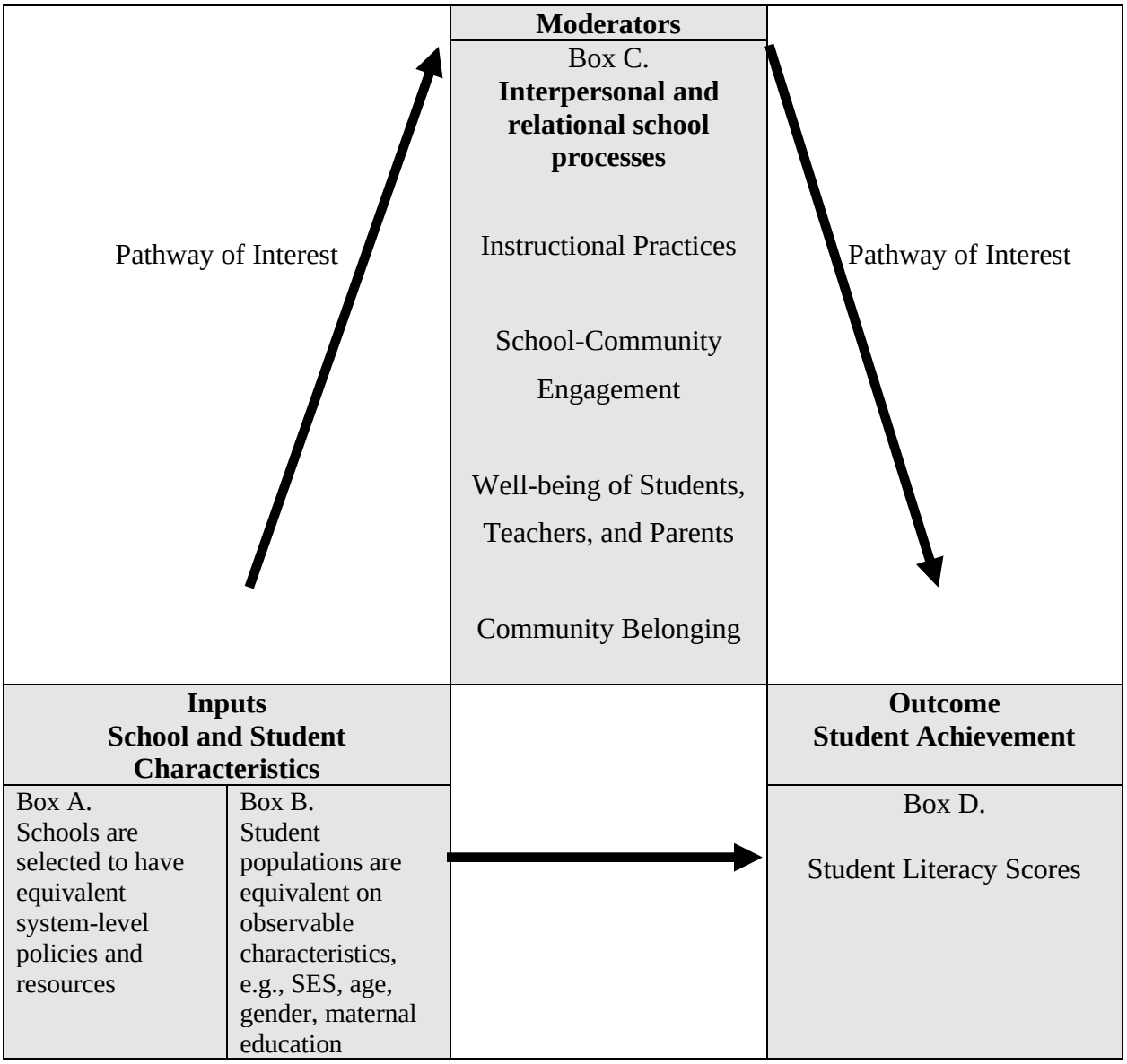

Figure 1. Conceptual framework: Levers for learning as moderators of student achievement

Instructional Practices

The remarkable progress globally in expanding access to education has enabled students from marginalized communities to attend school. Yet, these students typically enter school with less-developed school-relevant skills and less familiarity with school practices. Thus, how teachers relate to students' different needs within a classroom emerges as a particularly important relational dimension of instructional practice.

Recent impact evaluations of policies in low- and middle-income countries emphasize the concept of "teaching at the right level" (Banerjee et al, 2016). Banerjee et al. (2007) present causal evidence from a program that placed an extra teacher in classrooms to help teach lower-performing students, and Duflo et al. (2011) present causal evidence from splitting classrooms by students' baseline performance. Both papers show positive effects on student learning outcomes. However, other studies caution against tracking by ability (Oakes 1986; Steenbergen-Hu et al., 2016). Overall, an appropriate dosage of differentiated instruction, if adequately implemented, is hypothesized to be an important lever for literacy learning (Lesaux, Phillips Galloway, & Marietta, 2016).

At the same time, classroom-based studies in low- and middle-income countries consistently demonstrate that students' academic success is particularly at risk because of the predominance of teacher-centered pedagogy focused narrowly on students with high academic achievement (Schweisfurth, 2013; Vavrus and Bartlett, 2013; Ganimian, 2017). Centrally-mandated curriculum often also tailors content to high-skilled students (Banerjee and Duflo, 2012; Holla and Kremer, 2009). Under these conditions, teachers typically are able to increase learning for only a subset of students who are already likely to succeed in school.

Thus, based on the existing literature on instructional practices, we explored whether high- and low-performing schools varied in how teachers differentiate their instruction – either by meeting the needs of individual students (differentiated instruction) or by narrowly focusing their efforts only on a subgroup of skilled students (negative differentiation).

School-Community Engagement

Often, schools do not view families as a resource. Communication between schools and families can be difficult due to barriers of language, work schedules, or dissimilar expectations (Epstein, 2001; Glassman et al., 2007; Sanders, 2006; Uemura, 1999). Yet, evidence shows that when schools establish relationships and work together with communities and families, student learning improves (Kendall et al., 2015; Hong, 2011; Jeynes, 2003). In this study, we hypothesized that one lever for the higher average test scores in high-performing schools would be engaging with community, including involving parents in school decisions, and providing more opportunities for parent-teacher communication.

Several mechanisms may support the relationship between school-community engagement and higher learning outcomes, for instance, better coordination of services and closer understanding of parental needs (Epstein and Salinas, 1992); better access to information about schools that leads to stronger accountability (Andrabi, Das, & Khwaja, 2017; Mizala and Urquiola, 2013); and more opportunity for parental participation and engagement that in turn leads to schools that are responsive to parents' and students' needs (Bryk et al., 2010; World Bank, 2003).

Well-being: Students, Teachers, and Parents

Recent studies demonstrate that students and teachers from low-income households have physical and psychosocial challenges that stand in the way of teaching and learning (Banerjee and Duflo, 2012; Mullainathan and Shafir, 2013). In particular, students, teachers, and parents experience hunger, disease, lack of security, and mental stress, which are factors shown to negatively influence human capacities to function effectively (Behrman, 1996; Engle et al., 2007; Glewwe and Miguel, 2008; Grantham-McGregor et al., 2007; Taras, 2005; Walker et al.,

2007). When these types of survival challenges are overwhelming, mental capacity is devoted to thinking about these problems, leaving insufficient mental space for other activities such as teaching and learning. We predicted that in high-performing schools, students, parents and teachers would have higher levels of well-being than in low-performing schools.

Community Belonging

Students' sense of belonging to their schools and local communities can also impact their academic achievement. A sense of belonging to the school and the larger community includes respectful, collaborative community-building behaviors, in which individuals associate in groups to fulfill and protect their individual and collective interests and beliefs (Youniss et al., 2002). Schools can be spaces where young people develop the competencies to work cooperatively, negotiate differences, and practice the rights and obligations associated with being a member of a collective (Flanagan et al., 2011). Meta-analyses have demonstrated that when students have constructive interactions, they experience a sense of belonging to their schools and communities, and demonstrate higher achievement outcomes (Chapman et al., 2013; Osterman, 2000). Conversely, when students experience bullying or have negative social experiences with peers, their academic achievement can be compromised (Brown and Taylor, 2008 and Eriksen et al. 2012). We follow Eriksen et al.'s definition of bullying as "the exposure to repeated negative actions over time on the part of one or more students" (page 5). Social interactions in schools shape the social norms and attitudes of each student towards other students and teachers, and these relationships are of value to students as they grow up and learn (Carnoy & Marshall, 2005). We predicted that in schools where students reported a greater sense of community – including a

greater sense of school belonging, an understanding of how good citizens behave, and a lower frequency of being bullied – literacy outcomes would be stronger.

Research Design

Context

This analysis is part of the larger *Learning for All* project, conducted in Botswana, Brazil, Colombia, and Peru. In this paper, we focus on data collected in ten K-11 schools in Manizales, Colombia, a mid-sized city and the capital of one of the country's thirty-two departments. Its metropolitan area has approximately 550,000 inhabitants and 52 public K-11 schools. High-income families send their children to high-fee private schools; middle- and low-income families send their children to both public and low-fee private schools. Education in Colombia is highly decentralized, with municipalities (in this case, Manizales) making decisions about the allocation of resources; the highest education authority is the Secretary of Education, an appointed position by the mayor of the city. In collaboration with the Secretary and our local NGO partner, the Luker Foundation, we identified 10 schools serving similar populations. The Foundation is an NGO created by a local company interested in local research that aims to increase the social and human capital of the city. In partnership with the Foundation, data were collected between September and October 2015 in these ten schools. We surveyed teachers, parents, and students using instruments we developed for this study, drawing on a variety of validated measures, as detailed below and in a supplementary Appendix.[7]

[7] In addition to these measures, we also conducted semi-structured key informant interviews with school leaders; conducted classroom observations using the CLASS instrument; and engaged in in-depth participant observation and interviews with teachers, parents, and students, using an original "Day In the Life" observation and interview protocol; we do not present analyses of these data in this paper.

Validation of the research design: Student and household characteristics

Our sample included ten public schools in Manizales, similarly distributed by school type: five high-performing schools (HPS) and five low-performing schools (LPS), all serving the lowest-income households in the city (see Table 1). A cross-sectional sample of 1,486 students (464 in Grade 4, 531 in Grade 6, and 491 in Grade 8) clustered in 57 classrooms participated in this study. Our sample also included the parents (n=776) and teachers of the participating students (40 teachers across grades 4, 6, and 8).

We selected HPS and LPS based on publicly available data on academic achievement, averaged at the school level. Among the city's 30 public schools serving the lowest-income households, we chose the top and bottom five performers, as well as two back-up schools, based on the "Synthetic Index of Quality of Education" (Índice Sintético de Calidad Educativa). The index is created from nationally administered tests (*Saber*) and is used by the city government to rate schools on progress, performance, efficiency, and school climate.[8]

To validate our comparison, we verified whether the selected schools served similar populations. In Table 2, columns (1) and (2) display means by school type (high- or low-performing), and column (3) provides the mean difference between HPS and LPS.[9] We performed the following regression specification to estimate column (3):

$$X_{ics} = \gamma High_s + \varepsilon_{ics} \qquad (1)$$

[8] To use this index for our purposes, we averaged each school's ratings for primary and middle school, to cover our target populations of grades 4, 6, and 8. We used multiple imputations to impute missing values. From the roster of 30 schools, we excluded one school with not enough information to impute missing values, one school that specialized in serving disabled students, and one school focused on rural students. We sent our list of 10 selected plus two back-up schools to our NGO partner, who discussed with these school leaders the possibility of being part of this study. We used four of the selected high-performing and four of the low-performing schools and a back-up in each category (with one school uninterested in participation and another already participating in an intervention managed by our partner NGO), for a total of ten participating schools.

[9] Table 2 of the Appendix reports standard errors.

where X_{ics} is a student or household characteristic for student *i* in classroom *c* in school *s*, and $High_s$ is an indicator variable for high-performing schools. Standard errors are clustered at the classroom level. Note that equation (1) is estimated separately for each student and household characteristic.

Columns (4) through (9) present the same information further disaggregated by grade. For each set of variables (child outcomes, child characteristics, household characteristics), we report the p-value from a joint F-test with a null hypothesis that none of the variables in a given set are significantly different by school type (for definitions of all variables, see Appendix 1).[10]

Overall, student and household characteristics, obtained directly from our Student Survey, did not differ significantly by school type, with two exceptions. On average, students' age was similar in grades 4 and 6; however, in grade 8 student age differed significantly, with slightly younger students enrolled in HPS (grade 8 mean=14.88 years in HPS; mean=15.07 years in LPS). This difference may suggest that students were more likely to be held back in LPS. HPS tended to have more females than males (33.4 percent males, compared to 53 percent males in low-performing schools). As with student age, the sex difference increased with grade progression.

Household characteristics were measured using two sources of data: (1) the SISBEN (Sistema de Identificación de Potenciales Beneficiarios de Programas Sociales)[11], a household vulnerability index used in Colombia to channel social assistance (Castañeda, 2005) and (2) data

---

[10] We estimate regressions for each *T* variables:

$$y_{it} = x_{it}\beta_t + \varepsilon_{it} \qquad t = 1, \dots T$$

where y_{it} is the variable *t* of interest at the (*i*) student, household or classroom level, and x_{it} is the indicator variable for whether a school was in the low-performing group. We tested cross-equation restrictions such that $\beta_1 = \cdots = \beta_T = 0$. We used seemingly unrelated regression (SUR) models (Zellner, 1962). We applied the cluster-robust covariance estimator (Rogers, 1994).

[11] The SISBEN index is a means test used to identify poor households in Colombia. Level one and two (out of six) indicate the lowest-income households; most of the country's public health programs and social subsidies target these two strata.

from our Student Survey. The proportion of students coming from the lowest-income stratum (as indicated by SISBEN level one) was similar between HPS and LPS; overall, the majority of students were from the two lowest-income strata. Although not statistically significant, a higher proportion of students in HPS were from strata 3 or higher. Our survey data showed that indicators of socio-economic status were similar across the schools, as measured by an index of household assets, household size, and ownership of books. Regarding maternal education, the proportion with "university" level was higher in the HPS. However, for all other levels (primary, secondary, and technical degrees), the difference between HPS and LPS was not significant.

Teacher characteristics are displayed at the bottom of Table 2. Teachers in all schools were mostly female and had university degrees. Teachers in HPS were older (diff = 5.54 years), but we did not find significant differences in years of teaching experience by school type.

We conducted a joint test of differences for all variables in the student characteristics, the household characteristics, and the teacher characteristics sets. The results suggest that the two types of schools included in our sample served, on average, similar populations. The p-value from the joint *F*-test is 0.636 for student characteristics, 0.109 for household characteristics, and 0.324 for teacher characteristics.

Outcome Variable: Literacy Achievement

We assessed students' literacy in grades 4, 6, and 8 using the validated Spanish version of the Core Academic Language Skills Instrument (S-CALS-I), or the *Evaluación de Lenguaje Académico (ELA;* α =.86) (Uccelli et al, 2015; Meneses et al., 2018). The S-CALS-I is a 60-minute paper-and-pencil test designed to assess proficiency in the language for literacy in grades 4-8. The same full version of the S-CALS-I form was administered to all participating

students. It included a total of 68 items across nine tasks: understanding academic vocabulary, packing and unpacking complex words, ordering complex sentences, connecting ideas, tracking participants, interpreting writers' viewpoints, understanding metalinguistic terms, organizing analytic texts, and identifying definitions.[12] The total possible score was 65. We report standardized S-CALS-I scores.[13]

As expected, we found significant differences by school type for students' literacy performance as measured by the S-CALS-I. As shown in Table 2, students in HPS scored on average 0.432 standard deviations (SD) higher on the S-CALS-I relative to students in LPS. When we examined the outcome variable by grade, we generally observed widening test score gaps between HPS and LPS in upper grades; the test score gap was smaller in grade 4 (0.359 SD) than in grade 8 (0.602 SD). Figure 2 presents the distribution of S-CALS-I scores by type of school and by grade. Both the mean and the mode of the LPS were to the left of the HPS, and this pattern holds across grades.

Predictor Variables: Levers for Learning

Table 3 presents summary statistics of indices that measure the four hypothesized levers for learning. All indices were constructed using principal component analysis of their respective items on a polychoric correlation matrix and extracting the first component score. The indices are standardized (mean 0 and a standard deviation 1). Appendix 1 presents the detailed list of each index's components. We performed the following regression specification to estimate the difference between high- and low-performing schools:

[12] Most items in the S-CALS-I (or, in Spanish, ELA) are dichotomously scored as correct or incorrect. The partial-credit items were re-scaled to be between 0 and 1 so all items were equally weighted.
[13] We also collected data on a version of the PIRLS literacy test; we do not report the results here since PIRLS is not validated for individual observations; however, S-CALS and PIRLS scores were highly and positively correlated, and the results using S-CALS were very similar if PIRLS scores were used as the outcome measure.

$$Index_{ics} = \gamma High_s + \varepsilon_{ics} \quad (2)$$

where $Index_{ics}$ represents an index of one of the levers for learning, which are instructional practice (2 indices), school-community engagement (3 indices), well-being (3 indices), and community-belonging (3 indices), as described above. Standard errors were clustered at the classroom level. Equation (2) was estimated separately for each index variable.

Instructional Practices | Teacher Survey: On the basis of self-reported data gathered through the Teacher Survey, we generated two indices related to instructional practices: (1) the extent to which a teacher uses differentiated instruction, in which teachers adequately direct their efforts to meet the various needs of individual students (conceptualized as a feature of high-quality instruction); and (2) the extent to which a teacher engages in negative differentiation, in which teachers narrowly focus their efforts exclusively on a subgroup of skilled students (conceptualized as a feature of low-quality instruction). Our first index captures positive differentiated instruction by measuring teachers' self-reported tendency to tailor their pedagogy to individual students' level of ability, interests, and skills. This index uses items about the frequency of assigning different tasks to students who have difficulties; the frequency of assigning different work to students who move ahead faster; the frequency of assigning different tasks to students depending on their abilities; and the frequency of students' working in groups according to their abilities. Higher values on this index indicate higher levels of self-reported differentiated instruction.

Our second index of instructional practice captures negative differentiation. We measured the extent to which teachers self-report that they focus their efforts only on a subgroup of students who are better-prepared for school work, have higher skills, and demonstrate interest

in class, with less attention given to students who struggle in any of these areas. The former type of students are often characterized as “the front row” due to where they position themselves (or are positioned by teachers) in classroom spaces and their propensity to perform better or participate more (Ngware et al. 2013). This index comprises questions to teachers about their tendency to focus on a few students of greater ability; views of student diversity as a challenge to teaching; beliefs that not all students can learn; whether some students must be respected more than others; and the perception that (not) all students should have equal access to learning opportunities. A higher value on the index indicates greater levels of focus on higher-performing students.

Table 3 shows slightly higher but not significantly different levels of differentiated instruction between types of schools, with an average of 0.052 SD in HPS and 0.020 SD in LPS. However, we found significant differences in negative differentiation by school type. Teachers in HPS were more likely to focus their efforts only on a subgroup of skilled students. When we unpacked the negative differentiation index to examine item-level frequencies, we found that only 17 percent of teachers in high-performing schools agreed with the statement that students of all abilities can learn while 31 percent of teachers in low-performing schools agreed with the same statement. Moreover, 48 percent of teachers in HPS reported that they focused their teaching on higher-ability students in the class while only 30 percent of teachers in LPS reported doing so. These differences in negative differentiation between high- and low-performing schools were consistently large across all grades, with particularly large differences in grade 6 at 0.551 SD. However, we cannot reject the null hypothesis that the instructional practice lever was equivalent in high- and low-performing schools, given that the p-value from the joint F- test is 0.266.

School-Community Engagement | Teacher Survey and Teacher Time-Use Questionnaire: We measured school-community engagement using three indices. First, we constructed an index of school-community relationships by aggregating teachers' survey responses to whether their school cooperates with external organizations to conduct community-related activities (such as sports events, cultural and political activities, or activities for the benefit for disadvantaged groups). Second, we asked teachers about the extent to which their school provides parents or guardians with opportunities to actively participate in school decisions. Third, we collected detailed time-use questionnaires to measure the amount of time teachers spent communicating with their students' parents/guardians, over the course of their most recent workday.

Overall, we found similar levels of school-community engagement in HPS and LPS, as reported in Table 3 (p-value of 0.314 from joint F- test). For the indices of school-community relationships and teacher-parent communication, we cannot reject the null hypothesis that the measures are equivalent in HPS and LPS. In contrast, we found significant differences in parent participation. Parents whose children attend HPS reported lower levels of participation in school decisions compared to parents whose children attend LPS (difference = -0.549 SD). This pattern holds across grades, with particularly large differences for students in grade 6.

Student, Parent, and Teacher Well-Being | Student Survey, Parent Survey and Teacher Survey: We measured well-being through survey-based self-reports on levels of tiredness, sadness, stress, and irritation. Data were gathered through surveys administered independently to students, their teachers, and their parents. Descriptive statistics of student, teacher, and parent well-being measures are presented in Table 3. Overall, we found similar levels of well-being in HPS and

LPS as indicated by the results of the joint F- test (p-value = 0.667). It is worth noting, however, that students in grade 8 reported significantly higher levels of well-being in LPS (0.014 SD) than in HPS (-0.390 SD). This difference in student well-being is striking and intriguing given that younger students (grade 4 and 6) reported opposite patterns, with higher levels of well-being in HPS (though not statistically significant). Both parent well-being and teacher well-being indices were not significantly different between high- and low-performing schools.

Community Belonging | Student Survey: On the basis of data from the Student Survey, we measured belonging to their school community (in grades 4 and 6) and to the larger community beyond school (in grade 8). For fourth and sixth graders, we measured community belonging with adapted versions of the Pre-Adolescent Civic Engagement Scale (PACES) (Chi et al., 2006) and The Center for Information and Research on Civic Learning and Engagement's (CIRCLE) Indicators and Measures of Civic Outcomes (Nicotera et al., 2010). We generated two indices from these data. The first index measured students' sense of school belonging using items that asked whether a student felt out of place, like a stranger, or lonely at school. The second index measured the extent to which a student had been bullied at school over the past month, including questions about theft and psychological, emotional, or verbal abuse. For eighth graders, we adapted a measure previously established for this particular age-group in the International Civic and Citizenship Education Study (Schulz et al., 2011; Schulz et al., 2016). This index focused on good citizenship via fourteen items that measured students' understanding of requirements, expectations, and demands for a person to exhibit civic behavior (e.g., social respect and engagement, and compliance with rules and laws).

Table 3 presents descriptive statistics for the indices included in the lever on community belonging. Levels of school belonging are significantly higher in HPS (0.075 SD) compared to LPS (-0.084 SD), with the largest difference observed in grade 4 (0.325 SD). When we examined the items used to construct the school belonging index, we found that 37 percent of students reported feeling out of place and 32 percent of students report feeling alone in LPS. In contrast, 29 percent of students report feeling out of place and 25 percent of students report feeling alone in HPS.

Relatedly, levels of bullying were significantly lower in HPS (-0.133 SD) compared to LPS (0.147 SD).[14] The difference in bullying levels by school type was particularly large in grade 6 (-0.308 SD). The items used to construct the bullying index show that 50 percent of students in HPS reported ever being made fun of by classmates while over 63 percent of students in LPS reported such issues. Moreover, 62 percent in LPS reported having something stolen from them in school, in comparison to 50 percent in HPS.[15] For grade 8, HPS displayed higher levels of good citizenship (diff = 0.223 SD), although these differences were not statistically significant.

Overall, we reject the null hypothesis that community belonging is equivalent in high- and low-performing schools, given that the p-value from the joint F- test is 0.002.

Empirical Strategy

To analyze whether the index variables were associated with student literacy outcomes, we estimated three different versions of the following Ordinary Least Squares (OLS) regression:

[14] Our index of bullying include the following questions: 1. I was made fun of or called names; 2. I was left out of games or activities by other students; 3. Someone spread lies about me; 4. Something was stolen from me; 5. I was hit or hurt by other students; 6. I was made to do things I didn't want to do by other students

[15] Bullying affects both the recipients as well as the perpetrators. Unfortunately, we cannot measure the perpetrators of the bullying since the instrument did not identify who was/were the perpetrator(s).

$$y_{ics} = \alpha + \beta Index_{ics} + \gamma High_s + \rho Index_{ics} \times High_s + BX'_{ics} + \varepsilon_{ics} \quad (3)$$

where y_{ics} is the student learning outcome S-CALS-I for student i in classroom c in school s; $Index_{ics}$ represents an index of one of the levers for learning, which are instructional practice (2 indices), school-community engagement (3 indices), well-being (3 indices), and community-belonging (3 indices), as described above; and $High_s$ is an indicator variable for high-performing schools. X'_{ics} is a vector of student-level covariates, which includes grade, age, gender, household asset index, maternal education (dummies for each educational level), household size, and number of books in the home. The error term ε_{ics} consists of unobserved student ability or characteristics. Standard errors are clustered at the classroom level. Note that equation (3) was estimated separately for each of the 12 indices. As Equation (3) shows, we are exploiting variation at the school level (*s*), classroom (*c*) and individual (*i*) level.

First, we estimated a model only including $Index_{ics}$. The coefficient β captured the relation between student learning outcomes and the index of interest, controlling for student-level covariates. Second, we included (in addition to $Index_{ics}$) the variable $High_s$. We were interested in whether β changes with this inclusion; in other words, if the index of interest remained a significant predictor of learning outcomes even after accounting for whether the student attended a high- or low-performing school. Finally, we estimated the full model (Equation 1). The coefficient ρ captured whether the relation between student learning outcomes and the index of interest varied by school type. In the results presented in Tables 4 through 7, the column numbers under each index correspond to each of the three estimations.

Findings

Lever 1: Instructional Practice

We found that neither differentiated instruction nor negative differentiation were significant predictors of literacy outcomes in this sample of schools (see Table 4). We found no evidence that teachers' self-reported differentiated instruction was correlated differently with student literacy outcomes in HPS or LPS, as seen in Column (3) with a (close to) zero and non-significant coefficient for the index, as well as a negative and non-significant coefficient for the interaction of index and type of school. Similarly, we did not find any significant relation between student S-CALS-I scores and teachers' self-reported levels of negative differentiation.

Lever 2: School-Community Engagement

We did not find consistent evidence that our measures of school-community engagement predicted student literacy performance (see Table 5). For the index on school-community relationship, we did not find significant correlations between the index and S-CALS-I scores in the basic model (column 1) nor in the model that controls for school type (column 2). However, the interaction term of the index with school type was large and statistically significant. This interaction term suggests that the positive relation between a school's engagement with external community organizations and students' literacy performance was 0.162 SD larger in HPS, relative to LPS where the relation was indistinguishable from zero. In sum, even though levels of school-community relationship did not differ by school type, the positive contribution of this index to literacy outcomes was larger for HPS than for LPS in this sample.

Yet, we did not find consistent evidence that the relations between parents and schools – either through parent participation or time spent between teachers and parents – was a predictor of S-CALS-I scores or that this association was different for HPS and LPS.

Lever 3: Student, Parent, and Teacher Well-Being

We did not find systematic evidence of relationships between well-being measures and literacy outcomes (see Table 6). There was a positive, but not significant, association between student well-being and S-CALS-I. The association between parent well-being and S-CALS-I was positive and significant even after controlling for school type, yet it ceased to be significant once the interaction term was included. Indeed, both interaction coefficients for student well-being and parent well-being models were not significant, suggesting that these well-being indices were not differentially predicting literacy scores in high- and low-performing schools. For teacher well-being, we found a negative, but not significant, correlation with student literacy scores, and this association was similar across HPS and LPS.

Lever 4: Community Belonging

We found that sense of school belonging and bullying, measured for grades 4 and 6, were predictive of student literacy outcomes (see Table 7). The coefficient for school belonging was large, at 0.274 SD in the full model, and significant across all models. The interaction term was not significant, indicating that this relationship did not vary by school type. Yet, the level of school belonging was significant by type of school, indicating that even though the relationship did not vary by school type, school belonging in this sample was consistently higher in HPS.

For bullying, we found that a one-SD increase in a measure of student perception of bullying in schools was associated with a 0.099 SD decrease in literacy performance, holding constant school type, interaction of school type and index, and student-level covariates. The level of bullying was significantly higher in LPS. Finally, for grade 8, we found that good citizenship

was a significant predictor of S-CALS-I scores for the model with interactions: a one-SD increase in measures of students' understanding of civic behavior was correlated with a 0.151 SD increase in S-CALS-I scores. However, this interaction term was not statistically significant.

Selection issues

A limitation with our analyses is omitted variable bias since we were unable to account for all factors that may affect both levers for learning and literacy. To address this issue, we examined the stability of our estimates to omitted variable bias by performing the bounding approach proposed by Oster (2017) based on the framework in Altonji, Elder and Taber (2005). This method computes a lower bound coefficient estimate (β^*) by taking into account the changes in both coefficients and R-squared when control variables were included to the baseline regression model. The intuition behind the method is that if there is little change in R-squared, then the observed controls included in the model are not very informative. A key assumption of this method is that the relation between the coefficient of interest and unobservables can be recovered from the observed relation between the coefficient of interest and observable controls. To compute this lower bound coefficient, we assumed that the R-squared (R_{max}) was scaled up by a factor of $\Pi = 1.3$ as suggested in Oster (2017). We also assumed $\delta = 1$, which means that the unobservables were at least as informative as the observables. To compute the standard errors for β^*, we used a bootstrap approach.

Results of the regression from the bounding approach are presented in Appendix Table 2. The column "control" is our final model (the same as column 3 in Tables 4-7) and the column "adjusted" presents the lower bound coefficient estimates (β^*) for the final model. We found that our main results remained unchanged even when taking into account omitted variable bias.

First, the interaction of school-community relationship and school type was positive and statistically significant, indicating that the index was a significant predictor of literacy in high-performing schools but not in low-performing schools. Second, all of the indices for the community belonging lever remained statistically significant. Moreover, the absolute value of $\tilde{\delta}$ for the coefficient of each of the indices for the community belonging lever were all larger than 1 suggesting that unobservables were unlikely to change our core findings.[16] For example, in order to bring down the coefficient of bullying to zero, the unobservables would need to be 1.165 times more informative in predicting literacy than the full set of controls in our model.

Discussion

Below we discuss the two main findings that emerged from this study and close with directions for future research. Our analysis examined associations between interpersonal and relational dimensions of school processes and student literacy outcomes. We expanded on previous research that has focused on system-level policies (e.g., OECD, 2016) and on school inputs and student and family characteristics (Heyneman and Loxley, 1983b; Fuller, 1987; Baker, Goesling and LeTendre, 2002; Chiu & Khoo, 2005). While this study focuses on a single country and does not identify causal relationships, we aimed to widen the research lens of what is worth studying toward the goal of learning for all. Specifically, we hope that this study encourages new research on relational processes of what goes on regularly inside schools as potentially important, yet underinvestigated, levers for learning.

Finding 1: Variation Across Schools Serving Similar Student Populations

[16] A value of $|\tilde{\delta}| = 1$ means that the unobservables are at least as important as the observables included as controls. A negative $\tilde{\delta}$ means that if the observables are positively correlated with the index, the unobservables have to be negatively correlated with the index to yield a coefficient estimate of zero.

The schools included in the present sample were found originally using administrative data on national standardized tests and government data on school inputs and locations. To validate the study design, we applied our own literacy tests, with the expectation that our independent test would produce the same classification of schools. This validation confirmed that our sample included five pairs of high-performing and low-performing urban public schools, with each pair serving the same low-income neighborhood. We found that students in HPS displayed significantly higher literacy outcomes than those attending LPS in the grades investigated (i.e., 4, 6, and 8), even though the schools served similar low-income populations in the same communities. This finding is not trivial as there is limited empirical evidence that regular public schools in low- and middle-income countries are able to support meaningful learning for students in low-income communities. We cannot claim causality, given that we collected data as part of the daily reality of schools and not under an experimental design (e.g. we were not able to control for any unobservable differences across schools). Despite the limitations of the study design, we view the situated nature of our work as a strength that allows us to contextualize the functioning and constraints of the day-to-day lives of schools (National Research Council, 2003).

Relatedly, our literacy data illustrate that the literacy gap between students attending HPS and LPS is larger in eighth grade than in the upper elementary grades. Caution is needed because our data are cross-sectional and not longitudinal. We cannot rule out the possibility that this difference may be due to a cohort effect. With this caveat in mind, the observed pattern aligns with the widely discussed “Matthew effect” in literacy research, which refers to the widening gap between less and more skilled young readers as they develop over time (Stanovich, 1986). The observed correlation between gaps in literacy outcomes and grade level raises concern

about the possibility of accumulated disadvantages over time due to differences in instructional climate and opportunities between LPS and HPS (Lesaux et al., 2016).

Overall, this validation finding is consistent with prior studies that document that school-based measures contribute more to student achievement than family characteristics in low- and middle-income countries (Heyneman and Loxley, 1982, 1983a, 1983b; Fuller, 1987; Baker, Goesling and LeTendre, 2002; Chiu & Khoo, 2005). Yet our study, as discussed below, sought to advance the field by exploring if, even among schools that shared similar system-level policies, school inputs, and student characteristics, interpersonal and relational school processes would be associated with student achievement.

Finding 2: Community Belonging as a Lever for Learning

In testing four possible school-level levers for literacy outcomes, we found that only the indices in the community belonging lever differed significantly by school type and were significantly associated with literacy outcomes. The positive relation between school belonging and literacy outcomes was consistent across students in high- and low-performing schools. Importantly, students in low-performing schools reported a significantly lower level of school belonging, with a difference of 0.159 standard deviations. Bullying was also much more common in low-performing schools, with a difference of 0.280 standard deviations from high-performing schools. In addition, we found that in older students, understanding of how to be a good citizen is positively correlated with literacy outcomes, even after controlling for school type. These findings echo meta-analyses that point to the salience of students' sense of belonging and feeling part of a community, in explaining literacy outcomes (Chapman et al., 2013; Osterman, 2000; Brown and Taylor, 2008; Eriksen et al. 2012). Based on our results, it appears

that HPS were more likely than LPS to foster an environment in which students established and felt meaningful connections to their teachers and peers.

These results suggest that community belonging may be an important lever for learning. While we cannot establish causality, our findings on these two indices of community belonging suggest that additional research focused on this lever may shed light on the current global challenge of how to support pre-adolescents' literacy outcomes. A key area for future research is to examine if school-level interventions focused on developing a stronger sense of belonging, reducing bullying, and improving student understanding of civic behavior may lead to significant gains in student learning in low-income schools.

Directions for Future Research

The findings from this study point to the role of community belonging as a key lever for learning. However, we did not observe any significant associations between the three other theoretically important levers of learning – instructional practices, school-community engagement, and well-being – and student literacy outcomes. Below, we discuss directions for future research to better understand these other school processes.

For instructional practice, we foreground the need for further examination through direct observations of classroom practices and additional qualitative data (Gamoran 1989, Tomlinson 2014). For example, we found that practices of negative differentiation were more commonly reported in high-performing schools. To fully understand how and why teachers make instructional decisions and how students experience them requires complementary qualitative data.

For school-community engagement, we found that higher levels of cooperation

between schools and community groups were associated with higher literacy scores only among high performing schools. Given the unique conditions of urban public schools in low- and middle-income countries, there is a growing need to better understand how and why certain types of school-community engagement are practiced.[17] Perhaps schools engage differently with communities when the communities themselves present obstacles to learning, such as violence. Finally, vis-à-vis the well-being lever, we find that parents' well-being is the only positive correlate of literacy. However, this positive association between the index on parent well-being and literacy does not vary by school type.

In closing, we encourage more research on relational and interpersonal school processes in comparative education research, toward global goals of learning for all. This initial study offers a window into an understudied area and it provides an initial set of relevant research instruments; yet more studies and research tools are urgently needed to move this exploration further. Our findings point to community belonging as a promising lever, yet additional research on the other three hypothesized levers – instructional practices, community engagement, and well-being – is also needed. In particular, we call for interdisciplinary approaches, that combine economics, child development, and sociology, as well as for mixed-methods designs that integrate quantitative and in-depth qualitative analyses, as promising ways to uncover the complex nature of these interpersonal and relational school processes, and their contributions to learning.

References

[17] In low- and middle-income countries, a popular education reform has been school-based management, which decentralizes decision-making authority from governments to local communities. Schools included in our study did not have school-based management.

Altonji, J. G., Elder, T. E., & Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. *Journal of Political Economy*, *113*(1), 151-184.

Andrabi, T., Das, J., & Khwaja, A. I. (2017). Report cards: The impact of providing school and child test scores on educational markets. *American Economic Review*, *107*(6), 1535-63.

Baker, D. P., Goesling, B., & LeTendre, G. K. (2002). Socioeconomic status, school quality, and national economic development: A cross-national analysis of the "Heyneman-Loxley Effect" on mathematics and science achievement. *Comparative Education Review*, *46*(3), 291-312.

Banerjee, A. V., Cole, S., Duflo, E., & Linden, L. (2007). Remedying education: Evidence from two randomized experiments in India. *The Quarterly Journal of Economics*, *122*(3), 1235-1264.

Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukherji, S., Shotland, M. & Walton, M. (2016). Mainstreaming an effective intervention: Evidence from randomized evaluations of "Teaching at the Right Level" in India (No. w22746). *National Bureau of Economic Research*.

Banerjee, A. V. & Duflo, E. (2011). *Poor economics: A radical rethinking of the way to fight global poverty*. Public Affairs.

Behrman, J. R. (1996). The impact of health and nutrition on education. *The World Bank Research Observer*, *11*(1), 23-37.

Brown, S., & Taylor, K. (2008). Bullying, education and earnings: evidence from the National Child Development Study. *Economics of Education Review*, *27*(4), 387-401.

Bryk, A. S., Sebring, P. B., Allensworth, E., Easton, J. Q., & Luppescu, S. (2010). *Organizing Schools for Improvement: Lessons from Chicago*. University of Chicago Press

Carnoy, M., & Marshall, J. (2005). Cuba's academic performance in comparative perspective. *Comparative Education Review*, *49*(2), 230-261.

Carnoy, M., Ngware, M., & Oketch, M. (2015). The role of classroom resources and national educational context in student learning gains: comparing Botswana, Kenya, and South Africa. *Comparative Education Review*, *59*(2), 199-233.

Castañeda, T., (2005). Targeting social spending to the poor with proxy-means testing: Colombia's SISBEN system. *World Bank Social Protection Discussion Paper Series*, *0529*. World Bank, Washington D.C.

Chapman, R. L., Buckley, L., Sheehan, M., & Shochet, I. (2013). School-based programs for increasing connectedness and reducing risk behavior: A systematic review. *Educational Psychology Review*, *25*(1), 95-114.

Chi, B., Jastrzab, J., & Melchior, A. (2006). Developing Indicators and Measures of Civic Outcomes for Elementary School Students. *Center for Information and Research on Civic Learning and Engagement (CIRCLE), University of Maryland*, CIRCLE Working Paper 47.

Chiu, M. M., & Khoo, L. (2005). Effects of resources, inequality, and privilege bias on achievement: Country, school, and student level analyses. *American Educational Research Journal*, *42*(4), 575-603.

Coleman, J.S, et al. (1966) *Equality of Educational Opportunity*, U.S. Office of Education, Washington, D.C.

Duflo, E., Dupas, P., & Kremer, M. (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. *American Economic Review*, *101*(5), 1739-74.

Engle, P. L., Black, M. M., Behrman, J. R., de Mello, M. C., Gertler, P. J., Kapiriri, L., ... & Young, M. E. (2007). Child development in developing countries 3 Strategies to avoid the loss of developmental potential in more than 200 million children in the developing world. *The Lancet*, *369*, 229-42.

Epstein, J. L. (2001). *School, family, and community partnerships: Preparing educators and improving schools*. Westview Press.

Epstein, J. L., & Connors, L. J. (1992). School and family partnerships. *Practitioner*, *18*(4), n4.

Eriksen, T. L., Nielsen, H., & Simonsen, M. (2012). *The effects of bullying in elementary school* (No. 6718). Discussion Paper Series, Forschungsinstitut zur Zukunft der Arbeit.

Flanagan, C. A., M. L. Martinez, P. Cumsille, and T. Ngomane. (2011). Youth civic development: Theorizing a domain with evidence from different cultural contexts. In *Youth civic development: Work at the cutting edge. New Directions for Child and Adolescent Development*, edited by C. A. Flanagan and B. D Christens, 95-109.

Fuller, B. (1987). What school factors raise achievement in the third world? *Review of Educational Research*, *57*(3), 255-292.

Gamoran, A. (1989). Measuring Curriculum Differentiation. *American Journal of Education*, 97 (2):129–143.

Ganimian, A. (2017). Not drawn to scale? RCTs and education reform in developing countries. *RISE. Available at: https://www. riseprogramme. org/node/513 (Accessed 25 January 2019)*.

Glassman, D., Naidoo, J. P., Wood, F., Helmore, K., & O'Gara, C. (2007). *Community Schools in Africa*. New York: Springer Science & Business Media.

Glewwe, P., & Miguel, E. A. (2007). The impact of child health and nutrition on education in less developed countries. *Handbook of Development Economics*, *4*, 3561-3606.

Grantham-McGregor, S., Cheung, Y. B., Cueto, S., Glewwe, P., Richter, L., Strupp, B., & International Child Development Steering Group. (2007). Developmental potential in the first 5 years for children in developing countries. *The Lancet*, *369*(9555), 60-70.

Hanushek, E. A. (1979). Conceptual and empirical issues in the estimation of educational production functions. *Journal of Human Resources*, 351-388.

Hanushek, E. A. (2003). The failure of input-based schooling policies. *The Economic Journal*, *113*(485), F64-F98.

Heyneman, S. P., & Loxley, W. A. (1982). Influences on academic achievement across high and low income countries: A re-analysis of IEA data. *Sociology of Education*, 13-21.

Heyneman, S. P., & Loxley, W. A. (1983a). The distribution of primary school quality within high-and low-income countries. *Comparative Education Review*, *27*(1), 108-118.

Heyneman, S. P., & Loxley, W. A. (1983b). The effect of primary-school quality on academic achievement across twenty-nine high-and low-income countries. *American Journal of Sociology*, *88*(6), 1162-1194.

Holla, A., & Kremer, M. (2009). Pricing and Access: Lessons from Randomized Evaluations in Education and Health (Working Paper No. 158). Cambridge: Center for Global Development.

Hong, S. (2011). *A Cord of Three Strands: A New Approach to Parent Engagement in Schools*. Cambridge: Harvard Education Press.

Jeynes, W. H. (2003). A meta-analysis: The effects of parental involvement on minority children's academic achievement. *Education and Urban Society*, *35*(2), 202-218.

Kendall, N., Kaunda, Z., & Friedson-Rideneur, S. (2015). Community participation in international development education quality improvement efforts: current paradoxes and opportunities. *Educational Assessment, Evaluation and Accountability*, *27*(1), 65-83.L

Lesaux, N.K., Phillips Galloway, E. & Marietta, S.H. (2016). Leading Advanced Literacy Instruction in Linguistically Diverse Schools: A Guide for Education Leaders. New York: Guilford Press

Meneses, A., Uccelli, P., Santelices, M. V., Ruiz, M., Acevedo, D., & Figueroa, J. (2018). Academic Language as a Predictor of Reading Comprehension in Monolingual Spanish-Speaking Readers: Evidence From Chilean Early Adolescents. *Reading Research Quarterly*, *53*(2), 223-247.

Mizala, A., & Urquiola, M. (2013). School markets: The impact of information approximating schools' effectiveness. *Journal of Development Economics*, *103*, 313–335.

Mullainathan, S., & Shafir, E. (2013). *Scarcity: Why Having Too Little Means So Much*. New York: Times Books.

National Research Council. (2003). *Strategic education research partnership*. National Academies Press.

Nicotera, N., Altschul, I., Schneider-Muñoz, A., & Webman, B. (2010). Conceptual and Analytic Development of a Civic Engagement Scale for Pre-Adolescents. In J. Keshen, B. A. Holland, & B. E. Moely (Eds.), *Research for What? Making Engaged Scholarship Matter* (pp. 71–89). Charlotte, NC: Information Age Pub.

Oakes, J. (1986). Tracking, inequality, and the rhetoric of reform: Why schools don't change. *Journal of Education*, *168*(1), 60-80.

OECD (2012), *Equity and Quality in Education: Supporting Disadvantaged Students and Schools*, OECD Publishing.

OECD. (2016). *PISA 2015 results: Excellence and equity in education. Volume I*, OECD Publishing.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, *37*(2), 187-204.

Osterman, K (2000). Students' Need for Belonging in the School Community. *Review of Educational Research* 70 (3):323-67.

Põder, K., Lauri, T., & Veski, A. (2017). Does School Admission by Zoning Affect Educational Inequality? A Study of Family Background Effect in Estonia, Finland, and Sweden. *Scandinavian Journal of Educational Research*, *61*(6), 668-688.

Pritchett, L. (2013). *The rebirth of education: Schooling ain't learning*. Washington, DC: Center for Global Development.

Raitano, M., & Vona, F. (2013). Peer heterogeneity, school tracking and students' performances: evidence from PISA 2006. *Applied Economics*, *45*(32), 4516–4532.

Rogers, W. H. (1994). Comparison of nbreg and glm for negative binomial. *Stata Technical Bulletin*, *3*(16). Retrieved from https://ideas.repec.org/a/tsj/stbull/y1994v3i16sg16.4.html

Sanders, Mavis G. (2006). *Building school-community partnerships: collaboration for student success*. Thousand Oaks, Calif.: Corwin Press.

Schiefelbein, Ernesto, Joseph P. Farrell, and International Development Research Centre (Canada). 1982. *Eight years of their lives : through schooling to the labour market in Chile*. Ottawa, Canada: International Development Research Centre.

Schulz, W., Ainley, J., & Fraillon, J. (2011). *ICCS 2009 Technical Report* (Technical Report). Amsterdam: International Association for the Evaluation of Educational Achievement.

Schulz, W., Ainley, J., Fraillon, J., Losito, B., & Agrusti, G. (2016). *IEA International Civic and Citizenship Education Study 2016 Assessment Framework*. Amsterdam: Springer International Publishing.

Schweisfurth, M. (2013). Learner-centred education in international perspective. *Journal of International and Comparative Education*, *2*(1), 1-8.

Stanovich, Keith E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 22,* 360-407

Steenbergen-Hu, Saiying, Matthew C. Makel, and Paula Olszewski-Kubilius. (2016). What One Hundred Years of Research Says About the Effects of Ability Grouping and Acceleration on K–12 Students' Academic Achievement:Findings of Two Second-Order Meta-Analyses. *Review of Educational Research* 86 (4):849-899.

Taras, H. (2005). Nutrition and Student Performance at School. *Journal of School Health*, *75*(6), 199–213.

Todd, P. E., & Wolpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement. *The Economic Journal*, *113*(485).

Tomlinson, C.A. 2014. *The Differentiated Classroom: Responding to the Needs of All Learners*. Alexandria: ASCD

Toots, A., & Lauri, T. (2015). Institutional and contextual factors of quality in civic and citizenship education: Exploring possibilities of qualitative comparative analysis. *Comparative Education*, *51*(2), 247-275.

Uccelli, P., Barr, C. D., Dobbs, C. L., Galloway, E. P., Meneses, A., & Sanchez, E. (2015). Core academic language skills: An expanded operational construct and a novel instrument to chart school-relevant language proficiency in preadolescent and adolescent learners. *Applied Psycholinguistics*, *36*(5), 1077-1109.

Uemura, Mituse. (1999). Community Participation in Education: What Do We Know? Washington, D.C.: HDNED, World Bank.

UNICEF (Ed.). (2012). *Children in an urban world*. New York, NY: UNICEF.

Vavrus, Frances, and Lesley Bartlett. 2013. *Teaching in Tension: International Pedagogies, National Policies, and Teachers' Practices in Tanzania*. Rotterdam: Sense Publishers.

Walker, S. P., Wachs, T. D., Meeks Gardner, J., Lozoff, B., Wasserman, G. A., Pollitt, E., & Carter, J. A. (2007). Child development: risk factors for adverse outcomes in developing countries. *The Lancet*, *369*(9556), 145–157.

Wößmann, L. (2005). The effect heterogeneity of central examinations: evidence from TIMSS, TIMSS-Repeat and PISA. *Education Economics*, *13*(2), 143–169.

World Bank. (2003). *Making services work for poor people. World Development Report 2004*. Washington, D.C.: Oxford Univ. Press.

World Bank. (2015). World DataBank: Education Statistics Colombia. Washington, DC: World Bank.

World Bank. (2017). *Learning to Realize Education's Promise. World Development Report 2018*. Washington, D.C.: The World Bank.

Youniss, J., S. Bales, V. Christmas-Best, M. Diversi, M. McLaughlin, and R. Silbereisen. (2002). Youth Civic Engagement in the Twenty-First Century. *Journal of Research on Adolescence* 12 (1):121-148.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, *57*(298), 348–368.

Table 1. Summary of observations

	Number of units observed		
	High-performing	Low-performing	Total
School	5	5	10
Classrooms	28	29	57
Grade 4	*9*	*9*	*18*
Grade 6	*10*	*10*	*20*
Grade 8	*9*	*10*	*19*
Students & Parents	768	718	1486
Grade 4	*247*	*217*	*464*
Grade 6	*279*	*252*	*531*
Grade 8	*242*	*249*	*491*

Table 2. Summary statistics of student-level characteristics

	All			Grade 4			Grade 6			Grade 8		
	High	Low	Diff	High	Low	Diff	High	Low	Diff	High	Low	Diff
Child outcomes												
S-CALS-I (std)	0.207	-0.224	0.432***	0.169	-0.190	0.359*	0.155	-0.181	0.337**	0.307	-0.295	0.602***
Child characteristics												
Age (years)	12.75	13.02	-0.268***	10.51	10.72	-0.203	12.86	12.89	-0.030	14.88	15.07	-0.186*
Male (1=Yes)	0.334	0.529	-0.195	0.469	0.545	-0.075	0.312	0.538	-0.226	0.224	0.508	-0.284**
Joint F-test (p-value)			*0.636*			*0.396*			*0.229*			*0.444*
Household characteristics												
SISBEN strata												
One	0.292	0.352	-0.060	0.347	0.341	0.006	0.281	0.390	-0.110**	0.248	0.322	-0.075
Two	0.408	0.551	-0.143**	0.380	0.519	-0.139	0.388	0.524	-0.136	0.458	0.604	-0.146
Three +	0.301	0.097	0.204	0.273	0.139	0.133	0.331	0.085	0.246**	0.294	0.074	0.221
Index: Assets	0.097	-0.102	0.199	0.071	-0.185	0.256	0.001	-0.088	0.089	0.223	-0.046	0.269**
Household size	4.878	5.304	-0.426	5.034	5.174	-0.140	4.957	5.684	-0.727	4.630	5.042	-0.412
Mother's highest education												
Primary	0.184	0.318	-0.134	0.178	0.316	-0.138	0.192	0.280	-0.089	0.180	0.354	-0.174
Secondary (incomplete)	0.232	0.256	-0.024	0.178	0.265	-0.086**	0.265	0.234	0.032	0.237	0.271	-0.034
Secondary (complete)	0.359	0.294	0.064	0.368	0.219	0.148	0.331	0.350	-0.020	0.382	0.293	0.089
Advanced technical	0.077	0.052	0.025	0.052	0.052	0.000	0.082	0.0701	0.012	0.092	0.035	0.057
University	0.148	0.080	0.068*	0.224	0.148	0.076	0.131	0.065	0.065	0.110	0.048	0.062
Number of books at home												
Zero to two	0.231	0.289	-0.058	0.221	0.240	-0.019	0.254	0.327	-0.073*	0.220	0.288	-0.069
Three to ten	0.310	0.318	-0.008	0.279	0.291	-0.012	0.286	0.314	-0.028	0.374	0.346	0.027
More than ten	0.459	0.393	0.066	0.500	0.469	0.031*	0.459	0.358	0.101**	0.407	0.365	0.041
Joint F-test (p-value)			*0.109*			*0.424*			*0.212*			*0.113*
Teacher characteristics												
Female (1=Yes)	0.630	0.751	-0.121	0.547	0.885	-0.338	0.667	0.641	0.025	0.716	0.686	0.030
Age (Years)	48.36	42.82	5.54*	51.72	49.18	2.55	49.40	43.37	6.03	41.70	36.14	5.56
University degree (1=Yes)	0.930	0.956	-0.026	0.880	0.894	-0.014	1.000	1.000	0.000	1.000	1.000	0.000
Years in this school	5.618	8.699	-3.081	8.365	13.59	-5.221	2.192	8.290	-6.098*	5.292	4.880	0.413
Joint F-test (p-value)			*0.324*			*0.569*			*0.278*			*0.240*

p<0.01 ***; p<0.05 **; p<0.1 *

Note: Difference is the mean of high-performing schools minus the mean of low performing schools. Standard errors clustered at the classroom level. S-CALS-I (or

ELA) is standardized by grade. p-value from joint *F*-test examines whether the coefficients of school status (0=Low-performing / 1=High-performing) are jointly zero.

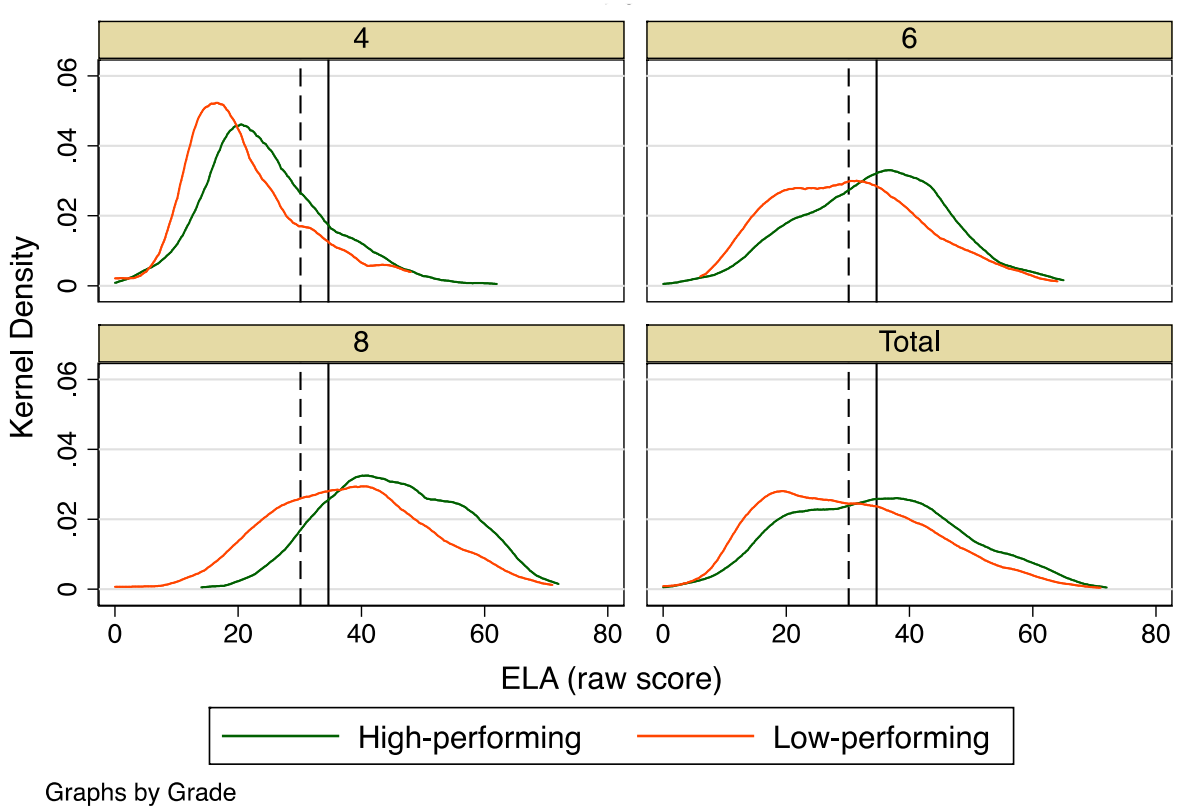

Figure 1. Distribution of literacy scores (S-CALS-I or ELA) by grade

Table 3. Summary statistics of school-level characteristics

	All			Grade 4			Grade 6			Grade 8		
	High	Low	Diff	High	Low	Diff	High	Low	Diff	High	Low	Diff
Instructional Practice												
Differentiated Instruction	0.052	0.020	0.032	-0.006	0.435	-0.441	0.189	-0.401	0.590	-0.050	-0.107	0.058
Negative Differentiation	-0.072	-0.513	0.441**	-0.180	-0.672	0.492	-0.090	-0.642	0.551**	0.098	-0.301	0.399
Joint F-test (p-value)			*0.266*			*0.053*			*0.938*			*0.719*
School-Community Engagement												
School-Community Relationship	-0.095	0.081	-0.175	0.004	-0.060	0.064	-0.058	0.774	-0.833	-0.255	-0.216	-0.039
Parent Participation	2.664	3.213	-0.549**	2.922	3.562	-0.640	2.184	3.000	-0.816**	2.943	3.032	-0.089
Teacher-Parent Communication	0.458	0.564	-0.106	0.335	0.866	-0.531*	0.435	0.580	-0.145	0.674	0.288	0.386
Joint F-test (p-value)			*0.314*			*0.492*			*0.372*			*0.692*
Well-being												
Student Well-being	-0.029	0.029	-0.058	0.159	0.070	0.090	0.125	0.012	0.113	-0.390	0.014	-0.404**
Parent Well-being	0.005	-0.007	0.012	0.082	0.159	-0.077	0.059	0.025	0.034	-0.202	-0.226	0.024
Teacher Well-being	0.168	0.335	-0.167	0.686	0.508	0.178	0.056	0.671	-0.615	-0.413	-0.012	-0.401
Joint F-test (p-value)			*0.667*			*0.221*			*0.225*			*0.032*
Community Belonging												
School Belonging	0.075	-0.084	0.159*	0.082	-0.244	0.325**	0.068	0.043	0.025	.	.	.
Bullying	-0.133	0.147	-0.280***	-0.199	0.040	-0.239	-0.074	0.234	-0.308**	.	.	.
Good Citizenship	0.097	-0.126	0.223	.	.	.	.	.	.	0.097	-0.126	0.223
Joint F-test (p-value)			*0.002*			*0.011*			*0.097*			*0.146*

p<0.01 ***; p<0.05 **; p<0.1 *

Note: Difference is the mean of high-performing schools minus the mean of low-performing schools. Standard errors clustered at the classroom level. p-value from joint *F*-test examines whether the coefficients of school status (0=Low-performing / 1=High-performing) are jointly zero.

Table 4. Analysis of literacy and Instructional Practice indices

Dependent variable:	S-CALS-I (standardized)					
Index:	Differentiated Instruction			Negative Differentiation		
	(1)	(2)	(3)	(1)	(2)	(3)
Index	-0.022	-0.019	-0.002	0.020	-0.082	-0.080
	(0.043)	(0.033)	(0.041)	(0.058)	(0.060)	(0.078)
High		0.365***	0.367***		0.407***	0.406***
		(0.094)	(0.093)		(0.105)	(0.099)
Index x High			-0.038			-0.004
			(0.074)			(0.125)
Constant	0.778**	0.572	0.583	0.619	0.431	0.432
	(0.371)	(0.360)	(0.360)	(0.389)	(0.378)	(0.390)
R2	0.146	0.175	0.176	0.146	0.178	0.178
Number of obs.	1,103	1,103	1,103	1,115	1,115	1,115
Number of clusters	45	45	45	45	45	45

p<0.01 ***; p<0.05 **; p<0.1 *

Note: Standard errors clustered at the classroom level. All regressions include student level covariates: grade, age, gender, asset index, mother's education, household size, and number of books at home.

Table 5. Analysis of literacy and School-Community Engagement indices

Dependent variable:	S-CALS-I (standardized)								
Index:	School-Community Relationship			Parent Participation			Teacher-Parent Communication		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Index	0.023	0.042	0.119	-0.095	-0.029	0.062	-0.141	-0.113*	-0.058
	(0.050)	(0.045)	(0.071)	(0.058)	(0.051)	(0.041)	(0.091)	(0.064)	(0.086)
High		0.383***	0.382***		0.351***	0.821***		0.345***	0.404***
		(0.095)	(0.090)		(0.100)	(0.274)		(0.085)	(0.117)
Index x High			0.162*			-0.157			-0.106
			(0.084)			(0.104)			(0.127)
Constant	0.459	0.301	0.224	0.921**	0.538	0.244	0.622	0.427	0.368
	(0.407)	(0.388)	(0.381)	(0.386)	(0.383)	(0.410)	(0.388)	(0.381)	(0.396)
R2	0.135	0.168	0.174	0.151	0.176	0.179	0.132	0.159	0.160
Number of obs.	1,038	1,038	1,038	1,115	1,115	1,115	1,163	1,163	1,163
Number of clusters	42	42	42	45	45	45	47	47	47

p<0.01 ***; p<0.05 **; p<0.1 *

Note: Standard errorsclustered at the classroom level. All regressions include the same student level covariates as in Table 4.

Table 6. Analysis of literacy and Student, Parent and Teacher Wellbeing indices

Dependent variable:	S-CALS-I (standardized)								
Index:	Student Wellbeing			Parent Wellbeing			Teacher Wellbeing		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Index	0.029	0.029	0.061	0.077**	0.076**	0.070	-0.066	-0.022	-0.032
	(0.030)	(0.029)	(0.048)	(0.033)	(0.033)	(0.054)	(0.055)	(0.052)	(0.064)
High		0.283***	0.284***		0.268**	0.268**		0.347***	0.342***
		(0.078)	(0.077)		(0.104)	(0.104)		(0.099)	(0.091)
Index x High			-0.057			0.011			0.021
			(0.059)			(0.066)			(0.090)
Constant	0.779**	0.592	0.606*	1.073**	0.818	0.820	0.594	0.404	0.405
	(0.356)	(0.356)	(0.359)	(0.532)	(0.541)	(0.539)	(0.382)	(0.375)	(0.375)
R2	0.120	0.139	0.140	0.135	0.151	0.151	0.141	0.167	0.167
Number of obs.	1,150	1,150	1,150	734	734	734	1,086	1,086	1,086
Number of clusters	57	57	57	56	56	56	44	44	44

p<0.01 ***; p<0.05 **; p<0.1 *

Note: Standard errors clustered at the classroom level. All regressions include the same student level covariates as in Table 4.

Table 7. Analysis of literacy and Community Belonging Indices

Dependent var.:	S-CALS-I (standardized)								
Index:	School Belonging			Bullying			Good Citizenship		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Index	0.230***	0.223***	0.274***	-0.091***	-0.075**	-0.099**	0.082	0.059	0.151*
	(0.037)	(0.036)	(0.043)	(0.030)	(0.031)	(0.040)	(0.054)	(0.055)	(0.084)
High		0.179*	0.176*		0.230**	0.229**		0.479***	0.483***
		(0.103)	(0.103)		(0.105)	(0.106)		(0.112)	(0.116)
Index x High			-0.095			0.043			-0.152
			(0.065)			(0.061)			(0.128)
Constant	0.334	0.209	0.174	0.319	0.168	0.172	0.569	0.554	0.554
	(0.448)	(0.457)	(0.449)	(0.474)	(0.478)	(0.479)	(1.630)	(1.588)	(1.595)
R2	0.160	0.167	0.170	0.117	0.130	0.130	0.201	0.244	0.248
Number of obs.	852	852	852	865	865	865	211	211	211
Number of clusters	38	38	38	38	38	38	19	19	19
Grade of students	4 & 6	4 & 6	4 & 6	4 & 6	4 & 6	4 & 6	8	8	8

p<0.01 ***; p<0.05 **; p<0.1 *

Note: Standard errors clustered at the classroom level. All regressions include the same student level covariates as Table 4.

Appendix 1. Definition of variables

Variable Name		Definition
Student-level variables		
SISBEN strata		National system of identifying beneficiaries for social subsidy. Classifies individuals into 6 strata according to socioeconomic level, stratum 1 is poorest and stratum 6 is richest.
Asset Index		First principal component **(FPC)** of household assets using: (has computer to do homework), (has functioning refrigerator or freezer) (has functioning DVD player). Mean 0, standard deviation **(SD)** 1. Higher values indicate higher socioeconomic status.
Instructional practice		
Differentiated Instruction		FPC of items: (frequency of assigning different jobs to student who have difficulties), (frequency of assigning different jobs to students who move faster), (frequency of assigning different jobs to students depending on their abilities), (frequency of student work in groups according to their abilities). Mean 0, SD 1. Higher value are interpreted as a proxy for positive classroom instructional practices.
Negative Differentiation		FPC of items: (some students must be respected more than others), (negative views of student diversity), (belief that not all students can learn essentials), (focuses on a few students who have the ability to learn) inverse of (*not* important that all student have equal access to learning). Mean 0, SD 1. Higher values are interpreted as a proxy for negative classroom instructional practices.
School-Community Engagement		
School-community relationship		FPC of items: Does the school cooperate with external groups/organizations in any of the following activities: (activities related to the environment, focused on the local area) (human rights projects) (activities related to disadvantaged people or groups) (cultural activities) (campaigns to raise people's awareness) (participating in sports events) Mean 0, SD 1. Higher values indicate more school-community engagement.
Parent participation		This school provides parents or guardians with opportunities to actively participate in school decisions.

		Scale of 1 (Strongly disagree), 2 (Disagree), 3 (Agree), to 4 (Strongly agree).
Teacher-parent communication		Hours that teachers spent with parents
Well-being		
Student Well-being		FPC of items: In the last month, have you experienced the following: (You woke up feeling tired) (You felt very sad, depressed, or down) (You had problems relaxing) (You felt easily irritated and annoyed) Mean 0, SD 1. Higher values indicate better well-being.
Parent Well-being		FPC of items: In the last month, have you experienced the following: (You woke up feeling tired) (You felt very sad, depressed, or down) (You had problems relaxing) (You felt easily irritated and annoyed) Mean 0, SD 1. Higher values indicated better well-being.
Teacher Well-being		FPC of items: In the last month, have you experienced the following: (You woke up feeling tired) (You felt very sad, depressed, or down) (You had problems relaxing) (You felt easily irritated and annoyed) Mean 0, SD 1. Higher values indicated better well-being.
Community Belonging		
School Belonging		FPC of items: (I feel like a stranger) (I feel weird and out of place) (I feel alone) Mean 0, SD 1. Higher values indicate stronger school belonging.
Bullying		FPC of items: (I was made fun of or called names) (I was left out of games or activities by other students) (Someone spread lies about me) (Something was stolen from me) (I was hit or hurt by other students) (I was made to do things I didn't want to do by other students) Mean 0, SD 1. Higher values indicate more bullying.

Good Citizenship		FPC of items: Belief that a good citizen is one who: (obeys the law) (votes in every election) (joins a political party) (works hard) (would participate in a peaceful protest against a law believed to be unjust) (knows about the country's history) (would be willing to serve in the military to defend the country) (follows political issues in the newspaper, on the radio, or on television) (participates in activities to benefit people in society) (shows respect for government leaders) (takes part in activities promoting human rights) (engages in political discussions) (takes part in activities to protect the environment) (is patriotic and loyal to the country) Mean 0, SD 1. Higher values indicate more civic engagement.

Table 2. Summary statistics of student-level characteristics

	All			Grade 4			Grade 6			Grade 8		
	High	Low	Diff	High	Low	Diff	High	Low	Diff	High	Low	Diff
Child outcomes												
S-CALS-I (std)	0.207	-0.224	0.432***	0.169	-0.190	0.359*	0.155	-0.181	0.337**	0.307	-0.295	0.602***
	(0.961)	(0.992)	(0.084)	(1.008)	(0.958)	(0.204)	(0.982)	(0.992)	(0.132)	(0.881)	(1.021)	(0.081)
Child characteristics												
Age (years)	12.750	13.020	-0.268***	10.510	10.720	-0.203	12.860	12.890	-0.030	14.88	15.07	-0.186*
	(2.012)	(2.024)	(0.003)	(0.843)	(0.934)	(0.050)	(1.098)	(1.064)	(0.013)	(1.035)	(1.062)	(0.018)
Male (1=Yes)	0.334	0.529	-0.195	0.469	0.545	-0.075	0.312	0.538	-0.226	0.224	0.508	-0.284**
	(0.472)	(0.499)	(0.039)	(0.500)	(0.499)	(0.082)	(0.464)	(0.500)	(0.098)	(0.418)	(0.501)	(0.019)
Joint F-test (p-value)			*0.636*			*0.396*			*0.229*			*0.444*
Household characteristics												
SISBEN strata												
One	0.292	0.352	-0.060	0.347	0.341	0.006	0.281	0.390	-0.110**	0.248	0.322	-0.075
	(0.455)	(0.478)	(0.033)	(0.477)	(0.475)	(0.081)	(0.450)	(0.489)	(0.003)	(0.433)	(0.468)	(0.023)
Two	0.408	0.551	-0.143**	0.380	0.519	-0.139	0.388	0.524	-0.136	0.458	0.604	-0.146
	(0.492)	(0.498)	(0.006)	(0.486)	(0.501)	(0.050)	(0.488)	(0.500)	(0.022)	(0.499)	(0.490)	(0.047)
Three +	0.301	0.0973	0.204	0.273	0.139	0.133	0.331	0.085	0.246**	0.294	0.0735	0.221
	(0.459)	(0.297)	(0.038)	(0.446)	(0.347)	(0.031)	(0.471)	(0.280)	(0.019)	(0.457)	(0.261)	(0.070)
Index: Assets	0.0967	-0.102	0.199	0.0709	-0.185	0.256	0.001	-0.088	0.089	0.223	-0.0459	0.269**
	(0.952)	(1.039)	(0.111)	(0.926)	(1.098)	(0.051)	(1.023)	(0.984)	(0.102)	(0.890)	(1.041)	(0.007)
Household size	4.878	5.304	-0.426	5.034	5.174	-0.140	4.957	5.684	-0.727	4.630	5.042	-0.412
	(2.027)	(3.668)	(0.163)	(2.279)	(2.366)	(0.064)	(2.115)	(5.035)	(0.553)	(1.594)	(2.870)	(0.121)
Mother's highest education												
Primary	0.184	0.318	-0.134	0.178	0.316	-0.138	0.192	0.280	-0.089	0.180	0.354	-0.174
	(0.388)	(0.466)	(0.037)	(0.384)	(0.466)	(0.164)	(0.395)	(0.450)	(0.014)	(0.385)	(0.479)	(0.028)
Secondary (incomplete)	0.232	0.256	-0.024	0.178	0.265	-0.086**	0.265	0.234	0.032	0.237	0.271	-0.034
	(0.422)	(0.437)	(0.049)	(0.384)	(0.443)	(0.004)	(0.442)	(0.424)	(0.044)	(0.426)	(0.445)	(0.081)
Secondary (complete)	0.359	0.294	0.064	0.368	0.219	0.148	0.331	0.350	-0.020	0.382	0.293	0.089
	(0.480)	(0.456)	(0.045)	(0.484)	(0.415)	(0.084)	(0.471)	(0.478)	(0.005)	(0.487)	(0.456)	(0.056)
Advanced technical	0.077	0.052	0.025	0.052	0.052	0.000	0.082	0.0701	0.012	0.092	0.035	0.057
	(0.267)	(0.222)	(0.031)	(0.222)	(0.222)	(0.042)	(0.274)	(0.256)	(0.043)	(0.290)	(0.184)	(0.010)
University	0.148	0.080	0.068*	0.224	0.148	0.076	0.131	0.065	0.065	0.110	0.048	0.062
	(0.356)	(0.272)	(0.009)	(0.418)	(0.357)	(0.043)	(0.338)	(0.248)	(0.011)	(0.313)	(0.214)	(0.014)
Number of books at home												

Zero to two	0.231	0.289	-0.058	0.221	0.240	-0.019	0.254	0.327	-0.073*	0.220	0.288	-0.069
	(0.422)	(0.454)	(0.032)	(0.416)	(0.428)	(0.026)	(0.437)	(0.470)	(0.011)	(0.415)	(0.454)	(0.081)
Three to ten	0.310	0.318	-0.008	0.279	0.291	-0.012	0.286	0.314	-0.028	0.374	0.346	0.027
	(0.463)	(0.466)	(0.051)	(0.449)	(0.455)	(0.021)	(0.453)	(0.465)	(0.005)	(0.485)	(0.477)	(0.146)
More than ten	0.459	0.393	0.066	0.500	0.469	0.031*	0.459	0.358	0.101**	0.407	0.365	0.041
	(0.499)	(0.489)	(0.019)	(0.501)	(0.500)	(0.005)	(0.500)	(0.481)	(0.007)	(0.493)	(0.483)	(0.064)
Joint F-test (p-value)			*0.109*			*0.424*			*0.212*			*0.113*
Teacher characteristics												
Female (1=Yes)	0.630	0.751	-0.121	0.547	0.885	-0.338	0.667	0.641	0.025	0.716	0.686	0.030
	(0.483)	(0.433)	(0.146)	(0.499)	(0.320)	(0.230)	(0.473)	(0.481)	(0.293)	(0.452)	(0.465)	(0.245)
Age (Years)	48.36	42.82	5.54*	51.72	49.18	2.55	49.40	43.37	6.03	41.70	36.14	5.56
	(9.97)	(10.83)	(3.30)	(10.51)	(10.76)	(5.48)	(7.68)	(11.95)	(5.76)	(8.14)	(4.15)	(4.17)
University degree	0.930	0.956	-0.026	0.880	0.894	-0.014	1.000	1.000	0.000	1.000	1.000	0.000
(1=Yes)	(0.255)	(0.205)	(0.062)	(0.326)	(0.309)	(0.157)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Years teaching in	5.618	8.699	-3.081	8.365	13.585	-5.221	2.192	8.290	-6.098*	5.292	4.880	0.413
this school	(5.022)	(9.625)	(2.164)	(5.21)	(12.40)	(4.695)	(2.203)	(9.569)	(3.197)	(3.886)	(2.430)	(1.996)
Joint F-test (p-value)			*0.324*			*0.569*			*0.278*			*0.240*

p<0.01 ***; p<0.05 **; p<0.1 *

Note: Difference is the mean of high-performing schools minus the mean of low performing schools. Standard deviation and standard errors in parentheses. Standard errors clustered at the classroom level. S-CALS-I (or ELA) is standardized by grade. p-value from joint *F*-test examines whether the coefficients of school status (0=Low-performing / 1=High-performing) are jointly zero.

Table 3. Summary statistics of school-level characteristics

	All			Grade 4			Grade 6			Grade 8		
	High	Low	Diff	High	Low	Diff	High	Low	Diff	High	Low	Diff
Instructional Practice												
Differentiated	0.052	0.020	0.032	-0.006	0.435	-0.441	0.189	-0.401	0.590	-0.050	-0.107	0.058
Instruction	(0.963)	(1.019)	(0.304)	(0.597)	(0.868)	(0.382)	(1.106)	(0.767)	(0.563)	(1.137)	(1.131)	(0.603)
Negative	-0.072	-0.513	0.441**	-0.180	-0.672	0.492	-0.090	-0.642	0.551**	0.098	-0.301	0.399
Differentiation	(0.707)	(0.600)	(0.209)	(0.578)	(0.547)	(0.298)	(0.435)	(0.311)	(0.213)	(1.046)	(0.701)	(0.540)
Joint F-test (p-value)			*0.266*			*0.053*			*0.938*			*0.719*
School-Community Engagement												
School-Community	-0.095	0.081	-0.175	0.004	-0.060	0.064	-0.058	0.774	-0.833	-0.255	-0.216	-0.039
Relationship	(1.065)	(0.940)	(0.323)	(0.845)	(0.740)	(0.452)	(1.165)	(0.292)	(0.537)	(1.171)	(1.108)	(0.625)
Parent	2.664	3.213	-0.549**	2.922	3.562	-0.640	2.184	3.000	-0.816**	2.943	3.032	-0.089
Participation	(0.890)	(0.673)	(0.247)	(0.959)	(0.497)	(0.426)	(0.807)	(0.000)	(0.308)	(0.583)	(0.856)	(0.384)
Teacher-Parent	0.458	0.564	-0.106	0.335	0.866	-0.531*	0.435	0.580	-0.145	0.674	0.288	0.386
Communication	(0.602)	(0.544)	(0.168)	(0.415)	(0.620)	(0.262)	(0.405)	(0.495)	(0.246)	(0.901)	(0.323)	(0.389)
Joint F-test (p-value)			*0.314*			*0.492*			*0.372*			*0.692*
Well-being												
Student Well-being	-0.029	0.029	-0.058	0.159	0.070	0.090	0.125	0.012	0.113	-0.390	0.014	-0.404**
	(1.059)	(0.938)	(0.098)	(0.993)	(0.934)	(0.109)	(1.038)	(0.979)	(0.152)	(1.063)	(0.902)	(0.152)
Parent Well-being	0.005	-0.007	0.012	0.082	0.159	-0.077	0.059	0.025	0.034	-0.202	-0.226	0.024
	(0.999)	(1.002)	(0.094)	(0.969)	(0.948)	(0.088)	(1.010)	(0.969)	(0.222)	(1.018)	(1.061)	(0.145)
Teacher Well-being	0.168	0.335	-0.167	0.686	0.508	0.178	0.056	0.671	-0.615	-0.413	-0.012	-0.401
	(0.830)	(0.716)	(0.239)	(0.432)	(0.576)	(0.256)	(0.725)	(0.504)	(0.380)	(0.919)	(0.781)	(0.469)
Joint F-test (p-value)			*0.667*			*0.221*			*0.225*			*0.032*
Community Belonging												
School Belonging	0.075	-0.084	0.159*	0.082	-0.244	0.325**	0.068	0.043	0.025			
	(0.995)	(1.000)	(0.090)	(1.004)	(1.048)	(0.135)	(0.989)	(0.943)	(0.111)			
Bullying	-0.133	0.147	-0.280***	-0.199	0.040	-0.239	-0.074	0.234	-0.308**	.	.	
	(0.999)	(0.981)	(0.091)	(0.964)	(0.999)	(0.138)	(1.028)	(0.961)	(0.108)	.	.	
Good Citizenship	0.097	-0.126	0.223	.	.	.	.	.	.	0.097	-0.126	0.223
	(1.025)	(0.957)	(0.153)	.	.	.	.	.	.	(1.025)	(0.957)	(0.153)
Joint F-test (p-value)			*0.002*			*0.011*			*0.097*			*0.146*

p<0.01 ***; p<0.05 **; p<0.1 *

Note: Difference is the mean of high-performing schools minus the mean of low-performing schools. Standard deviation and standard errors in parentheses. Standard errors clustered at the classroom level. p-value from joint *F*-test examines whether the coefficients of school status (0=Low-performing /

1=High-performing) are jointly zero.

Appendix Table 4. Bounding approach estimation

Lever:	**Instructional Practice**			
Index:	Differentiated Instruction		Negative Differentiation	
	Control	Adjusted	Control	Adjusted
Index	-0.002	-0.054	-0.08	-0.202
	(0.041)	(0.082)	(0.078)	(0.121)
High	0.367***	0.317***	0.406***	0.354
	(0.093)	(0.072)	(0.099)	(0.103)
Index x High	-0.038	-0.126	-0.004	0.220
	(0.074)	(0.164)	(0.125)	(0.220)
R-sq	0.176	0.233	0.178	0.237
$\tilde{\delta}$ for $\beta_{Index} = 0$		3.904		-1.715
$\tilde{\delta}$ for $\beta_{High} = 0$		4.043		2.386
$\tilde{\delta}$ for $\beta_{High\ x\ Index} = 0$		-1.312		0.035

Lever:	**School-Community Engagement**					
Index:	School-Community Relationship		Parent Participation		Teacher-Parent Communication	
	Control	Adjusted	Control	Adjusted	Control	Adjusted
Index	0.119	0.213	0.062	0.057	-0.058	-0.175
	(0.071)	(0.075)	(0.041)	(0.128)	(0.086)	(0.112)
High	0.382***	0.340***	0.821***	1.063	0.404***	0.298
	(0.090)	(0.063)	(0.274)	(3.079)	(0.117)	(0.277)
Index x High	0.162*	0.246***	-0.157	-0.150	-0.106	-0.804
	(0.084)	(0.094)	(0.104)	(0.691)	(0.127)	(0.322)
R-sq	0.174	0.231	0.179	0.238	0.16	0.212
$\tilde{\delta}$ for $\beta_{Index} = 0$		-3.299		1.532		2.570
$\tilde{\delta}$ for $\beta_{High} = 0$		4.597		0.276		1.350
$\tilde{\delta}$ for $\beta_{High\ x\ Index} = 0$		24.646		-0.158		-0.255

Lever:	**Well-being**					
Index:	Student Wellbeing		Parent Wellbeing		Teacher Wellbeing	
	Control	Adjusted	Control	Adjusted	Control	Adjusted
Index	0.061	0.023	0.070	0.078	-0.032	-0.027
	(0.048)	(0.046)	(0.054)	(0.064)	(0.064)	(0.156)
High	0.284***	0.220***	0.268**	0.216***	0.342***	0.287***
	(0.077)	(0.074)	(0.104)	(0.080)	(0.091)	(0.073)
Index x High	-0.057	-0.047	0.011	0.070	0.021	-0.101
	(0.059)	(0.073)	(0.066)	(0.111)	(0.09)	(0.210)
R-sq	0.14	0.186	0.151	0.200	0.167	0.222
$\tilde{\delta}$ for $\beta_{Index} = 0$		0.189		2.931		0.430
$\tilde{\delta}$ for $\beta_{High} = 0$		3.177		3.771		3.231
$\tilde{\delta}$ for $\beta_{High\ x\ Index} = 0$		2.468		-0.268		0.252

Lever:	**Community Belonging**					
Index:	School Belonging		Bullying		Good Citizenship	
	Control	Adjusted	Control	Adjusted	Control	Adjusted
Index	0.274***	0.428***	-0.099**	-0.116*	0.151*	0.082*
	(0.043)	(0.116)	(0.040)	(0.059)	(0.084)	(0.051)
High	0.176*	0.122	0.229**	0.179**	0.483***	0.330*
	(0.103)	(0.081)	(0.106)	(0.076)	(0.116)	(0.201)
Index x High	-0.095	-0.097	0.043	0.040	-0.152	-0.048
	(0.065)	(0.115)	(0.061)	(0.102)	(0.128)	(0.543)
R-sq	0.17	0.226	0.13	0.173	0.248	0.330
$\tilde{\delta}$ for $\beta_{Index} = 0$		1.147		1.165		1.480
$\tilde{\delta}$ for $\beta_{High} = 0$		2.864		3.495		2.046
$\tilde{\delta}$ for $\beta_{High\ x\ Index} = 0$		0.573		0.604		1.275

p<0.01 ***; p<0.05 **; p<0.1 *

Note: Standard errors clustered at the classroom level. All regressions include student level covariates: grade, age, gender, asset index, mother's education, household size, and number of books at home.