

Early Schooling: Teachers Make the Difference

School enrollment in the early grades is close to universal in Latin America and the Caribbean, but the quality of education is generally poor. As a result, many children in the region learn little in their first years of formal schooling. Since early education is considered vital for economic and social progress, the region's failure in this area is of great concern.

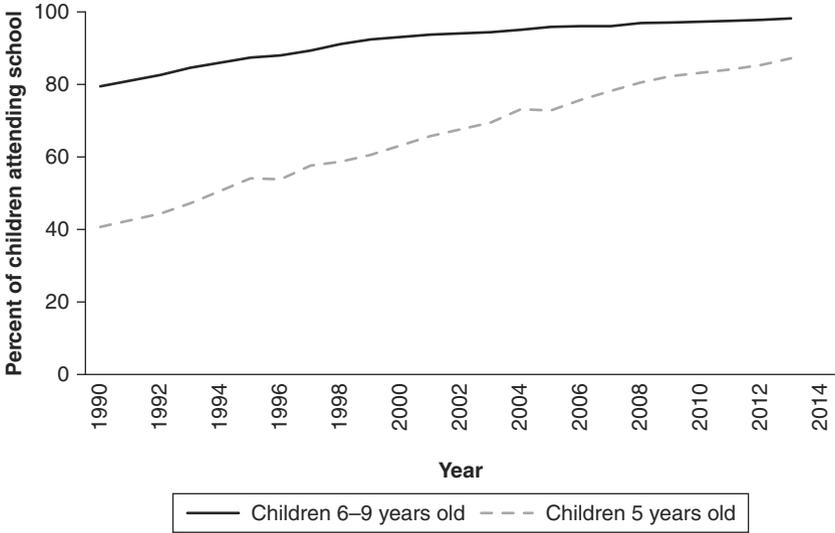
The lack of cognitive (and other) skills of Latin American and Caribbean workers is frequently cited as a major reason behind the low economic growth rates of the region (Hanushek and Woessmann 2012). Policymakers are generally aware of the poor performance of secondary school students from Latin America on the Program for International Student Assessment (PISA) tests. However, skill formation is a cumulative process. It is hard to learn in late childhood and adolescence without a solid foundation. Fixing the problem of low quality in secondary school may do little for learning outcomes if the same is not done for the quality of schooling in the early years.

Early School Enrollment: A Regional Success

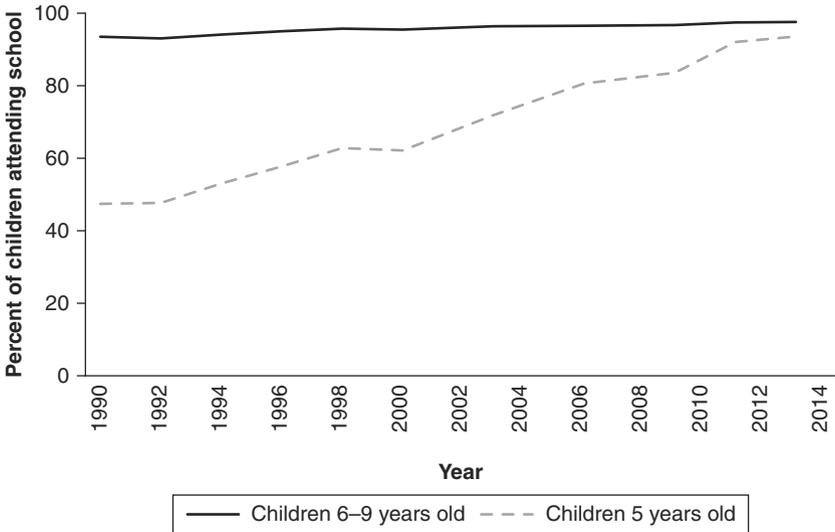
By and large, young children in Latin America and the Caribbean attend school. Figure 5.1 plots the proportion of children enrolled in school between 1990 and 2014 in six countries: Brazil, Chile, Honduras, Jamaica, Mexico, and Panama.¹ Changes in these countries are broadly representative of those that have taken place throughout the region. The solid line focuses on children aged 6–9.

Figure 5.1 School Attendance, 1990–2014

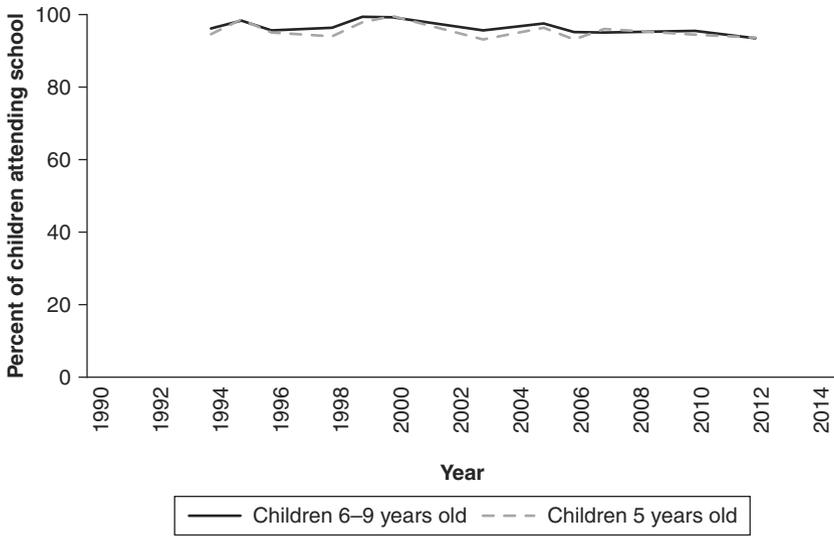
a. Brazil



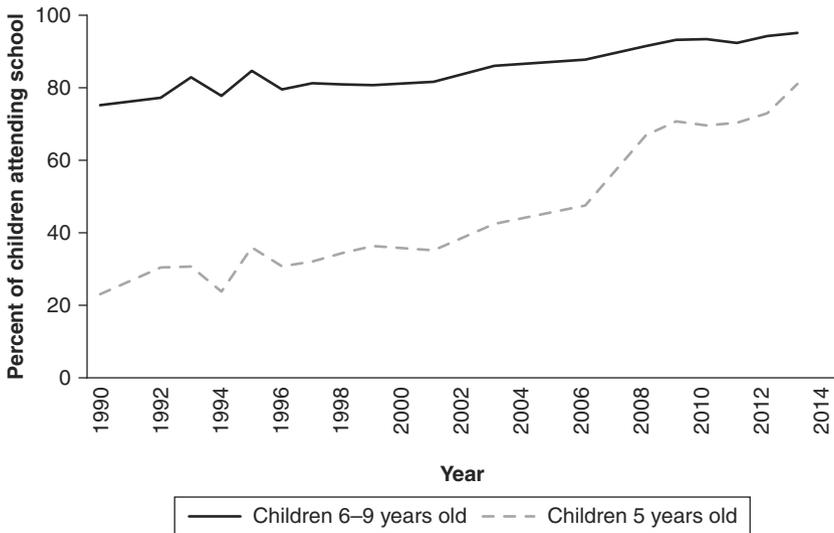
b. Chile



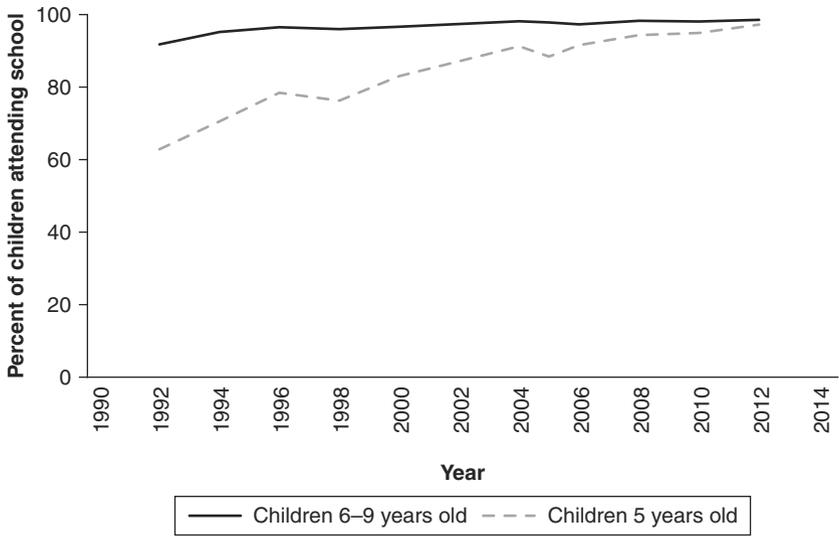
c. Jamaica



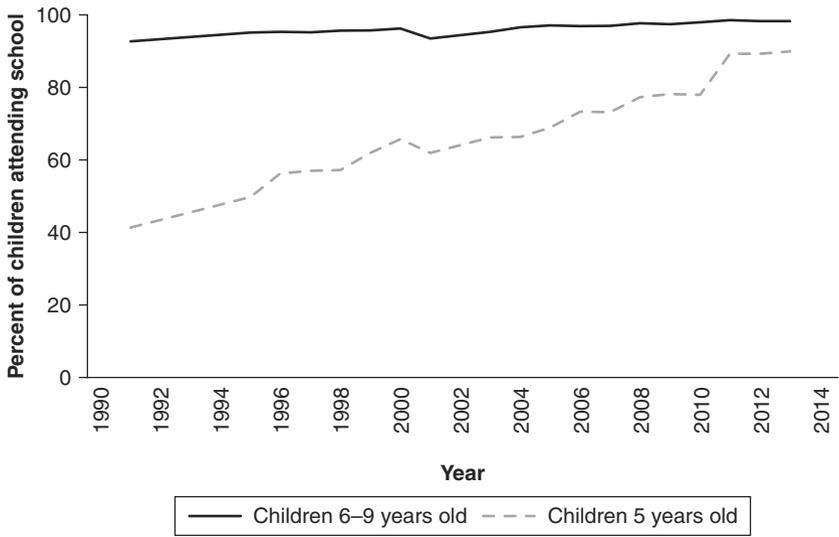
d. Honduras



e. Mexico



f. Panama



Source: Authors' calculations based on Inter-American Development Bank Harmonized Household Surveys.

For children who start school on time and do not repeat grades, this corresponds to enrollment in roughly first through third or fourth grade of primary school, depending on the exact birth date.² In virtually every country in the region, including the poorest ones, school attendance for children aged 6–9 years is universal or very close to universal. In many countries, this has been the case since at least the early 2000s.

Good evidence from the region, particularly from the Southern Cone countries, shows that children who attend pre-primary school are better prepared to learn in later grades. Argentina implemented a large program to expand the coverage of pre-primary education between 1993 and 1999. Children who benefited from this expansion in coverage performed better on mathematics and Spanish achievement tests in third grade. One year of pre-primary school increased children's test scores by 0.23 standard deviations. Children who attended pre-primary school also demonstrated superior participation skills, such as attention, effort, class participation, and discipline, as reported by their teachers (Berlinski, Galiani, and Gertler 2009).

In Uruguay, pre-primary school attendance for children 4 and 5 years old resulted in a significant, positive effect on the number of years of schooling completed. By age 15, children who had attended some pre-primary school were 27 percent more likely to be in school than their peers who did not, and had on average completed 0.8 more years of schooling (Berlinski, Galiani, and Manacorda 2008).³

The dashed line in Figure 5.1 shows that many countries in the region have made a sharp push to extend the coverage of education for 5-year-olds.⁴ Moreover, wealth gradients in enrollment have declined substantially. For example, in Chile, the difference in the proportion of 5-year-olds attending school between the first and fifth wealth quintiles fell from 29 to 5 percentage points between 2000 and 2013; in the Dominican Republic, it fell from 33 to 7 percentage points; and in Panama it fell from 45 to 15 percentage points (Table 5.1). Increasingly, the enrollment gap in kindergarten between the rich and the poor is closing, much as it did in the first grades of primary school in earlier decades.

Table 5.1 School Attendance, by Wealth Quintile

Country	Difference between richest and poorest quintiles (percentage points)			
	5-year-old children		6- to 9-year-old children	
	2000	2013	2000	2013
Argentina	16	5	2	0
Bolivia	30	31	9	1
Brazil	35	13	9	2
Chile	29	5	4	0
Colombia	33	13	9	1
Costa Rica	36	28	4	2
Dominican Republic	33	7	7	5
Ecuador	21	8	5	1
El Salvador	52	35	25	7
Honduras	42	26	20	4
Mexico	21	4	5	2
Panama	45	14	4	2
Paraguay	23	30	9	2
Peru	36	12	3	3
Uruguay	16	5	2	0

Note: When data for 2000 or 2013 were not available, data from the nearest year were used. For 2000, data from 2001 were used for Brazil, Costa Rica, Honduras, Nicaragua, and Paraguay. For 2013, data from 2012 were used for Mexico and Nicaragua. Data for Uruguay correspond to urban areas only.

Source: Authors' calculations based on Inter-American Development Bank Harmonized Household Surveys.

A Low Score in Learning Outcomes

Although young children in Latin America and the Caribbean go to school, many of them appear to learn very little. Data on test scores in the first grades of pre-primary and primary school are scarce, and (as with child development) comparability is an issue. Only two countries in the region—Chile and Honduras—participated in the 2011 application of the Trends in International Mathematics and Science Study (TIMSS) of fourth grade students, and only two countries—Colombia and Honduras—participated in the 2011 application of the Progress in International Reading Literacy Study (PIRLS) of fourth grade students.

It is not possible to credibly benchmark early learning outcomes of children in Latin America and the Caribbean relative to other countries outside the region. However, a large number of countries in Latin America have participated in a regional test of language and math applied to third graders in Latin America in 2007 (Segundo Estudio Regional Comparativo y Explicativo, or SERCE) and 2013 (Tercer Estudio Regional Comparativo y Explicativo, or TERCE). These data can therefore be used to compare the learning outcomes of young children in different countries in the region.

There are substantial differences in math test scores across countries in 2013, as is shown in Table 5.2.⁵ Chile, Costa Rica, Uruguay, and Mexico performed relatively well; the Dominican Republic, Nicaragua, Paraguay, and Panama did not. Third graders in Chile, the top performer, posted scores that were 0.9 standard deviations higher, on average, than children in the Dominican Republic, the country with the lowest average scores.⁶

SERCE and TERCE are, by construction, comparable in their difficulty. This makes it possible to analyze changes in a country's

Table 5.2 Math Test Scores, by Country and Year

	SERCE (2007)	TERCE (2013)
Argentina	0.03	0.22
Brazil	0.03	0.27
Chile	0.19	0.55
Colombia	-0.01	0.13
Costa Rica	0.25	0.39
Ecuador	-0.18	0.16
Guatemala	-0.29	0.01
Mexico	0.21	0.33
Nicaragua	-0.18	-0.10
Panama	-0.25	-0.04
Paraguay	-0.09	-0.08
Peru	-0.17	0.22
Dominican Republic	-0.69	-0.35
Uruguay	0.26	0.34

Notes: All scores have been redefined as standard deviation units of the 2007 score. SERCE = Segundo Estudio Regional Comparativo y Explicativo; TERCE = Tercer Estudio Regional Comparativo y Explicativo.

Source: Authors' calculations based on data from SERCE and TERCE.

performance over time. Test scores have increased in all countries, but the magnitude of these changes varies widely. In Peru, Chile, the Dominican Republic, and Ecuador, test scores have increased by 0.3 standard deviations or more. In Paraguay, Nicaragua, Uruguay, Mexico, Costa Rica, and Colombia, improvements have been more modest: 0.15 standard deviations or less.

Countries also vary in the proportion of the total variation in TERCE test scores that is accounted for by differences across and within schools. The cross-school component explains almost half of the variability in test scores in Panama, Paraguay, and Peru. Conversely, the cross-school component is much smaller—about one-quarter—in Chile, Costa Rica, and Mexico. This finding suggests that policies to raise the learning outcomes of the lowest-performing students that target specific schools, rather than children within schools, are more likely to be effective in some countries (such as Peru) than in others (such as Chile). Box 5.1 presents a methodological discussion of how these cross- and within-school components can be estimated.

Box 5.1 The Variance in Test Scores: Is It the Child or the School?

From a policy point of view, it is important to know whether the variation in test scores in early primary school in a country results primarily from the fact that some schools have lower average scores than others or, rather, from differences between children within the same school. A decomposition of the variance in test scores into across-school and within-school components attempts to answer this question. Such a calculation was carried out using the 2013 TERCE (Tercer Estudio Regional y Comparativo), a regional test applied to third graders in Latin America in 2013.

In its simplest form, this decomposition can be calculated by a regression of test scores in a country on school fixed effects. The R-squared in this regression measures the between-school variance in test scores. The R-squared in these regressions was about 0.5 in Panama, Paraguay, and Peru, but only about 0.25 in Chile, Costa Rica, and Mexico. However, sampling error may be a problem for these “naïve” estimates of the across- and within-school components of the variance. This is because the number of schools and of children tested per school varied a great deal across countries in TERCE.

The robustness of the results from the basic decomposition (in particular, the relative ranking of countries) to the presence of sampling error was tested in two ways. In one approach, a new sample was created. By construction, this new sample had exactly the same number of schools and the same number of children per school, in each country. Specifically, schools with eight or fewer tested children (the value at the 25th percentile for the sample as a whole) were discarded, and a sample of 158 schools (the number of schools in the country with the fewest schools in the sample, Colombia) and 9 children per school was randomly chosen in every country. (The sample for each country has exactly 1,422 children.) One hundred iterations of this procedure were carried out, regressions of test scores on school fixed effects were run in each sample, and the average R-squared for these 100 iterations was calculated. Using this procedure, Peru, Panama, and Colombia (rather than Paraguay) were estimated to have the highest across-school variance, while Chile, Costa Rica, and Mexico were estimated to have the lowest across-school variance, as before.

The second approach was inspired by the literature on teacher value added (see Box 5.2). In this literature, it is standard to calculate the variance of teacher effects. It has long been recognized, however, that sampling error tends to overstate the true variance of teacher effects. Specifically, $V_o = V_t + V_e$, where V_o is the observed variance, V_t is the true variance, and V_e is the variance of the measurement error. One approach to correcting the observed variance is to estimate the variance of the measurement error using an Empirical Bayes procedure.⁷ The same approach can be used to correct the variance of school (rather than teacher) effects for sampling error in TERCE. When the variance is uncorrected, the three countries where the school effects explain the largest fraction of the total variance in test scores are Paraguay, Peru, and Honduras; these are also the countries in which differences across schools explain the largest fraction of the total variance after the Empirical Bayes correction. Similarly, when the variance is uncorrected, the three countries where the school effects explain the smallest fraction of the total variance in test scores are Chile, Costa Rica, and Mexico; these are also the countries in which differences across schools explain the smallest fraction of the total variance after the Empirical Bayes correction.

In sum, no matter how the variance is decomposed, there are some countries, like Peru, where a great deal of the variation in child test scores is driven by differences across schools, and others, like Chile, where much more of this variation is driven by differences across children in the same school.

Data on the socioeconomic characteristics of children and their parents collected in SERCE and TERCE are limited and are missing for many children. Better data to study wealth gradients in learning outcomes are available in three country-specific tests: the Cerrando Brechas study of kindergarten students in 2012 in Ecuador, the Exámenes de la Calidad y el Logro Educativos (EXCALE) of first graders in 2011 in Mexico, and the Evaluación Censal de Estudiantes (ECE) of second graders in 2010 in Peru.

To analyze socioeconomic gradients in math scores, children in Cerrando Brechas, ECE, and EXCALE were sorted into quintiles of the national distribution of wealth in each country.⁸ These results, reported in Table 5.3, show that the richest children scored 0.5 standard deviations higher than the poorest children in Ecuador and Mexico, and a full standard deviation in Peru.⁹ Table 5.3 also shows that in Ecuador and Peru boys have higher scores than girls, while the reverse is true in Mexico.

Table 5.3 Wealth Gradients in Math Scores

	Cerrando Brechas (Kindergarten, Ecuador)	EXCALE (first grade, Mexico)	ECE (second grade, Peru)
Second quintile	0.069 (0.042)	0.070 (0.064)	0.100*** (0.018)
Third quintile	0.159*** (0.052)	0.202*** (0.063)	0.458*** (0.022)
Fourth quintile	0.296*** (0.055)	0.384*** (0.060)	0.712*** (0.023)
Fifth quintile	0.532*** (0.070)	0.524*** (0.045)	0.996*** (0.031)
Girl	-0.036** (0.017)	0.061* (0.032)	-0.057*** (0.013)
R-squared	0.022	0.063	0.110
Observations	14,243	6,776	60,646

Notes: Coefficients and standard errors (in parenthesis). Units are in standard deviations. Children in the first (poorest) wealth quintile are the omitted category. Standard errors clustered at the school level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Source: Authors' calculations based on Cerrando Brechas test of kindergarten students in 2013 in Ecuador, the Exámenes de la Calidad y el Logro Educativos (EXCALE) of first graders in 2011 in Mexico, and the data for the Evaluación Censal de Estudiantes (ECE) of second graders in 2010 in Peru.

It is also possible to estimate socioeconomic gradients in early learning outcomes in Jamaica. Samms-Vaughan (2005) analyzes the evolution of test scores using a (relatively small) sample of approximately 250 children; these children were followed from pre-K to third grade. Samms-Vaughan compares the performance of children on the Reading, Spelling and Arithmetic subscales of the Wide Range Achievement Test (WRAT) in households that are asset-rich and asset-poor. At pre-K age, the mean difference between those in the high- and low-asset groups was between 0.6 and 0.8 standard deviations; by third grade, socioeconomic gradients had widened considerably, to between 1.0 and 1.3 standard deviations.¹⁰

In sum, the evidence makes clear that learning outcomes in Latin America and the Caribbean are poor, especially in some countries, and among children from households that are poorer and where parents have less schooling.

Grading Classroom Quality

The fact that children are enrolled in school but many of them appear to learn very little suggests that the quality of early schooling is a serious problem in the region.¹¹

Classroom quality is a complex, multifaceted construct, but (much as is the case with the quality of daycare, discussed in Chapter 4) it can be separated into two discernible components: structural and process quality. Structural quality focuses on features of the classroom experience such as the environment, the nature and level of teacher training and experience, adoption of certain curricula, class size, and student-teacher ratios. Process quality, on the other hand, refers to a student's direct interactions with resources and opportunities in the classroom. This includes the ways teachers implement lessons, the nature and quality of interactions between adults and students or between students and their peers, and the availability of certain types of activities.

Structural Quality: A Lesser Factor

Some studies in the United States have found that lower child-to-teacher ratios in the early grades improve child learning

outcomes. The best known of these studies is Project STAR in Tennessee, an intervention that randomly assigned children in kindergarten through third grade to “small” (13–17 students per teacher) or “large” classes (22–25 students). Children in the smaller classes outperformed those in the large classes in the short run (Krueger 1999), although some of these effects faded out as children became older (Krueger and Whitmore 2001). Remarkably, Chetty and others (2011) find that children randomly assigned to smaller classes outperform those in larger classes on a number of measures of adult performance, including college attendance, roughly two decades later.

Nevertheless, these results do not appear to be the norm. Reviews of hundreds of interventions in the United States conclude that the evidence that these structural features have a direct impact on children’s academic achievement or social development is mixed (Hanushek 2003; NICHD Early Child Care Research Network 2002).

These findings have been corroborated with research on developing countries. Murnane and Ganimian (2014) review 115 well-designed impact evaluations of educational interventions in over 30 lower- and middle-income countries, and conclude that learning outcomes were not consistently improved by better materials, classroom technology, flexible education funding grants, or smaller class sizes, unless the day-to-day interactions of children and teachers were also targeted. Kremer, Brannem and Glennerster (2013) arrive at a similar conclusion.¹²

Process Quality: The Real Test

When adults are sensitive and responsive to children’s cues and needs, children begin to learn and develop (National Scientific Council 2012). As children enter formal schooling, interactions among teachers and students in the classroom begin to play a critical role in development.

How students spend time: A key component of process quality is how students spend their classroom time. A growing body of literature indicates that the extent to which students are engaged in educationally focused activities in the classroom predicts academic and

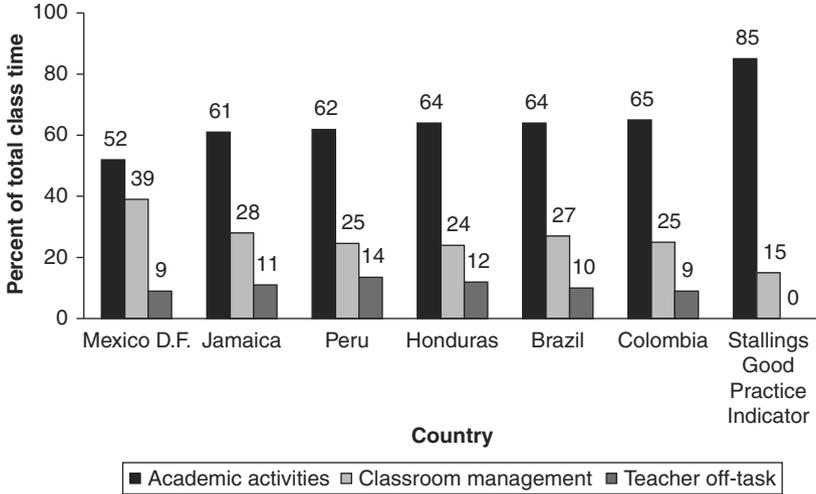
social outcomes. Specifically, when instruction is targeted to a specific skill, that particular skill is developed (NRP 2000; Snow, Burns, and Griffin 1998). For example, language and literacy instruction is related to greater skill gains in language and literacy (Piastra and others 2012), while greater emphasis on math and science is associated with greater skills gains in math and science (Clements and Sarama 2011; Sarama and Clements 2009). Within the social and emotional domain, teachers who model explicitly and teach about emotions help students develop knowledge about emotions and regulation (Denham, Bassett, and Zinsler 2012).

Several large-scale studies in the United States have carefully studied how time is spent in the classroom (Early and others 2005; Hamre and others 2006; La Paro and others 2009). These studies generally conclude that a substantial amount of time in most classrooms is spent on noninstructional activities such as routines and transitions.

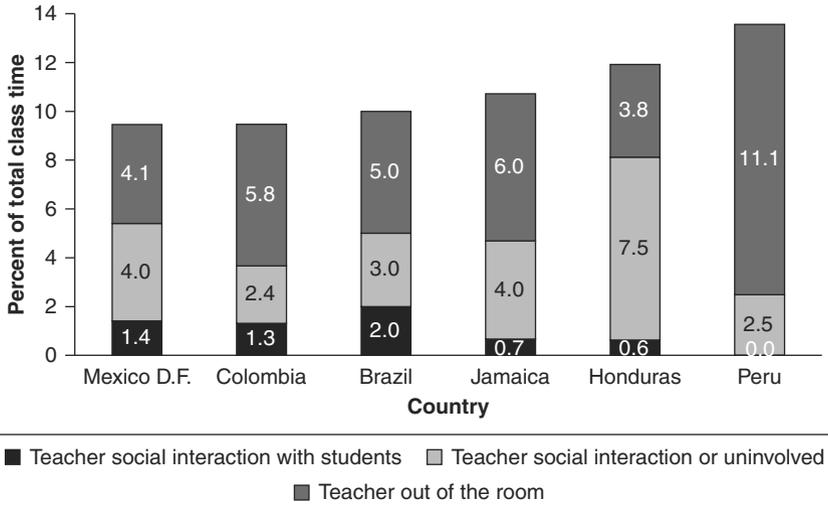
Results from Latin America and the Caribbean present a similar picture. Bruns and Luque (2015) report the results from the application of the Stallings Classroom Snapshot instrument (Stallings 1977) in more than 15,000 classrooms in six countries in Latin America and the Caribbean (Brazil, Colombia, Honduras, Jamaica, Mexico, and Peru).¹³ Figure 5.2 reproduces some important results from their research. On average, only between 50 and 65 percent of time in the classroom in the six countries is spent on instruction, well below the Stallings good practice benchmark of 85 percent. This means that, even in the best-performing countries in the region, a full day of instruction is lost per week, relative to the good practice benchmark. In every country in the region, between 8 and 14 percent of time is lost because teachers are physically absent from the classroom altogether (e.g., arriving late or leaving early), or are engaged in social interactions with other adults (e.g., chatting at the classroom door). Thus, in a 200-day school year, students on average miss 20 full days of instruction. Even when a teacher is spending time on instruction, it is comparatively rare for all of the students in the classroom to be engaged; more often than not, more than half the children are not paying attention and are disengaged or bored (Bruns and Luque 2015).

Figure 5.2 How Teachers Spend Their Time in the Classroom

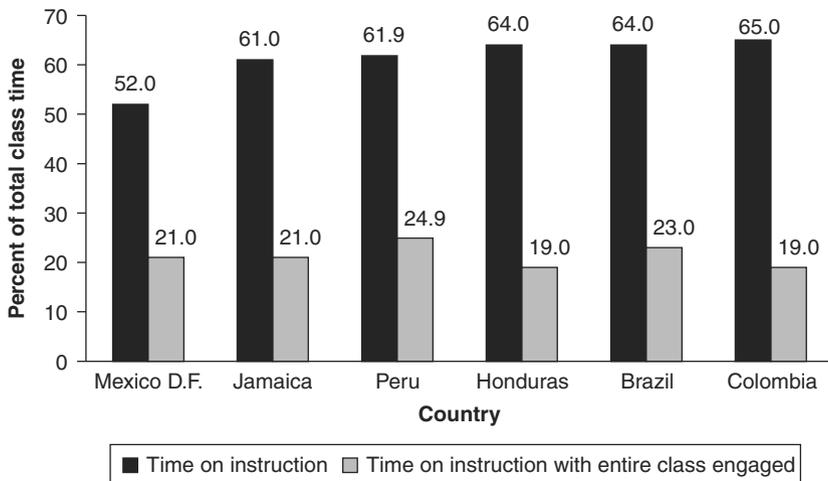
a. Proportion of Time Spent on Different Classroom Activities



b. A Breakdown of Teacher Time Off-Task



c. Time on Instruction and Time on Instruction with Entire Class Engaged



Source: Bruns and Luque (2015).

The quality of teacher-child interactions: Over the past two decades, researchers in the United States have focused on another key aspect of students' classroom experience—the elements of teachers' interactions that promote positive development. This body of work considers the moment-to-moment learning experiences students have with their teacher and peers. A number of studies have found that the quality of students' interactions with one another and with their teachers is more important for their achievement on evaluations of academic preparedness than aspects of structural quality.¹⁴

Given this growing body of evidence that the quality of teacher-child interactions matter for students' development, more recent work has focused on articulating a clear framework of what effective interactions look like. Building on previous descriptions of quality teaching (Brophy and Good 1986; Eccles and Roeser 2005), the Teaching Through Interactions framework (Hamre and Pianta 2007) has become a widely used and accepted model for understanding and measuring the quality of teacher-child interactions. It is aligned with a classroom observation tool, the Classroom Assessment Scoring System (CLASS) (Pianta, La Paro, and Hamre 2008b), used in much of the research in this area (see Box 4.1). The Teaching Through

Interactions framework describes three domains of interactions that have strong theoretical and empirical backing: emotional support, classroom organization, and instructional support.

Emotional support: In classrooms with high levels of emotional support, teachers and students have positive relationships and enjoy spending time together. Teachers are aware of, and responsive to, children's needs, and prioritize interactions that place an emphasis on students' interests, motivations, and points of view. In classrooms with low levels of emotional support, teachers and students appear emotionally distant from one another, and there are instances of frustration in interactions. Teachers seldom attend to children's need for additional support and, overall, the classroom follows a teacher's agenda with little opportunity for student input. Many studies from the United States have found associations between the teachers' provision of emotionally supportive interactions in the classroom and students' social-emotional development.¹⁵

Classroom organization: In highly organized classrooms, teachers are proactive in managing behavior by setting clear expectations; classroom routines allow for students to get the most out of their time engaged in meaningful activities; and teachers actively promote students' engagement in those activities. In less organized classrooms, teachers might spend much of their time reacting to behavior problems; classroom routines are not evident; students spend time wandering or not engaged in activities; and teachers do little to change this. When teachers manage behavior and attention proactively, students spend more time on-task and are better able to regulate their attention (Rimm-Kaufman and others 2009). Students in better organized and managed classrooms also show larger increases in cognitive and academic development (Downer and others 2010).¹⁶

Instructional support: In classrooms with high levels of instructional support, a teacher promotes higher order thinking and provides quality feedback to extend students' learning. At the low end, rote and fact-based activities might be common, and students receive little to no feedback about their work beyond whether or not it is correct. In these classrooms, teachers do most of the talking or the room is quiet. The quality of instructional support provided in a

classroom is most consistently linked with higher gains in academic outcomes, such as test scores.¹⁷

Although the three domains of teacher-student interactions are conceptually distinct, and can be measured separately, it is frequently observed that teachers who excel in one domain also excel in the other two. For this reason, when taken together, the behaviors associated with higher levels of emotional support, classroom organization, and instructional support can be described as “Responsive Teaching” (Hamre and others 2014).

Though students learn and develop more in classrooms with higher quality teacher-child interactions, few students actually experience these types of interactions in the early years of school. For example, in the United States, emotional support and classroom organization are typically of only moderate quality in primary classrooms, and instructional support is moderate to low.¹⁸ A recent study using data from Finland finds qualitatively similar results (Salminen 2013).

There is a small, but growing body of evidence from Latin America on teaching practices in the early grades, with a focus on the interactions between teachers and students. Cruz-Aguayo and others (2015) report the results from the application of two instruments that measure different aspects of classroom quality in a sample of 78 kindergarten, first-grade, and second-grade classrooms in three countries in Latin America: Brazil, Chile, and Ecuador. The first instrument is the SNAPSHOT (Ritchie and others 2001), which (like the Stallings) focuses on the activity that is happening within a classroom at a given moment.¹⁹ The second instrument that was used is the CLASS. In addition to these two instruments, data were collected on some aspects of structural quality, including student-teacher ratios.

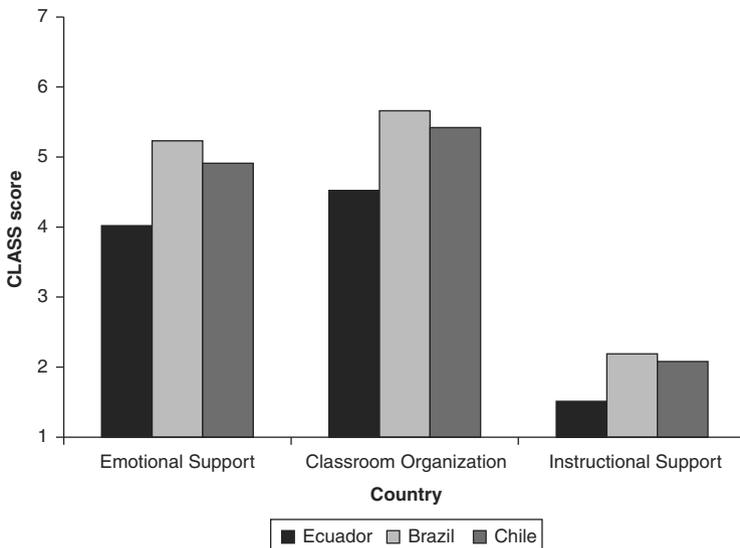
The results from the SNAPSHOT suggest that the bulk of instructional time in all three countries involved students working on a whole group activity (e.g., students sitting at their desks and copying teacher-provided sentences from the chalk- or white-board, or reproducing letters) or individually (e.g., students working on identical pages in their workbooks). Typically, children sat at small desks, with all desks facing the front of the classroom. Students rarely worked collaboratively in small groups. In some classrooms in all

three countries, no activity occurred during much of the time, and students were provided little or no direction.

The results from the CLASS, summarized in Figure 5.3, indicate that in all three countries, scores on the Emotional Support and Classroom Organization domains of the CLASS were in the mid-range, while scores on Instructional Support were consistently very low, especially (but not only) in Ecuador.²⁰ Similar patterns were found in a nationally representative sample of kindergarten classrooms in Ecuador (as reported in Araujo and others 2014). Like others (Bruns and Luque 2015), Cruz-Aguayo and others (2015) also emphasize that students' classroom experiences frequently were very different across classrooms within the same school.

Leyva and others (2015) assess time use and the quality of teacher-student interactions, as measured by the CLASS, and the association between these dimensions of process quality and student outcomes, in a sample of 91 public prekindergarten classrooms in Chile. A substantial amount of time was spent on noninstructional activities, such as eating snacks, transitions, and recess.²¹ Higher

Figure 5.3 The Classroom Assessment Scoring System Domain Scores, Brazil, Chile, and Ecuador



Source: Cruz-Aguayo and others (2015).

levels of instructional support were associated with larger gains on one of the three tests of language and early literacy, and one of the two tests of executive function.

One of the best-designed studies on the effects of teacher quality and the classroom environment on child learning outcomes is Araujo and others (2014), who study a sample of 454 public kindergarten classrooms in the coastal area of Ecuador. The paper begins by discussing the difficulty of credibly identifying the effects of better teachers, or better teaching practices if students are assigned to teachers on the basis of unobserved characteristics. The study avoids this problem by randomly assigning an entering cohort of approximately 15,000 kindergarten students in 202 schools, all of which had at least two kindergarten classes, to teachers. Compliance with the random assignment was very high at 98 percent.

During the school year, Araujo and others (2014) collected very rich data on teachers and students. They used the CLASS to measure the quality of interactions between teachers and students in the classroom. They also collected additional information on teachers, including whether they were tenured or worked on a contract basis; their years of experience in the teaching profession; their intelligence, as measured by the Wechsler Adult Intelligence Scales (WAIS-IV); their personality, as measured by the Big Five personality inventory; teacher executive function; and a variety of characteristics about teachers and their environments when they were children. At the end of the year, they collected 12 separate tests of child learning, including four tests each of early literacy, early math, and executive function.

Based on this careful design, Araujo and others (2014) first show that a child randomly assigned to a teacher at the 95th percentile of the distribution of quality learns, on average, 0.18 standard deviations more than a child assigned to an average teacher. In terms of magnitude, this effect is comparable to that found in a randomized evaluation of a program that gave cash transfers equivalent to about 10 percent of household expenditures to extremely poor households in Ecuador (0.18 standard deviations, as reported by Paxson and Schady [2010]), and to estimates of the impact of a year of pre-primary education in Argentina (0.23 standard deviations, as reported by Berlinski, Galiani, and Gertler [2009]).²² The effects are

also similar to those reported in the literature on teacher effects in the United States (as summarized in Hanushek and Rivkin [2012]).

Moreover, the same teachers improve child learning outcomes in all domains (language, math, executive function) and these teachers improve the learning outcomes of all children in a classroom (those with higher and lower baseline levels of development, those whose parents have more or less education, and those in households with higher or lower levels of wealth) by roughly the same amount. As the authors put it “a rising tide” (in this case, random assignment to a better teacher) “lifts all boats” (in this case, improves the learning outcomes of all children).

To what extent do different teacher characteristics or behaviors affect child learning in kindergarten? Much as is the case in the United States, very inexperienced teachers (teachers with three or fewer years of teaching experience) produce less learning: a child randomly assigned to a “rookie” teacher, defined in this way, learns on average 0.16 standard deviations less than a child assigned to a more experienced teacher. (Above three years, there are no returns to experience in terms of child learning outcomes.) Whether a child is assigned to a tenured teacher or to one who works on a contract basis does not affect her learning gains. Smarter teachers, as measured by the WAIS-IV, produce more learning, but the effects are small.²³ None of the personality traits measured by the Big Five (neuroticism, extraversion, openness, agreeableness, and conscientiousness) is significantly associated with more or less student learning. Similarly, a teacher’s score on a test of executive function and various measures of her early environment (such as her parents’ education) do not predict student learning. On the other hand, teacher-student interactions, as measured by the CLASS, are strongly predictive of child learning outcomes (Araujo and others 2014).

Lessons for Policy

In most countries in Latin America and the Caribbean, enrollment in the first grades of primary school is essentially universal, and the proportion of children who go to kindergarten (including children from poor households) is rising rapidly. It is not unreasonable to

assume that, much as happened with enrollment in primary school, enrollment in kindergarten will no longer be a challenge for most countries in the region in the foreseeable future.

What is a challenge for the region is quality—particularly process quality. There is clear, consistent evidence that the quality of the education many children receive is poor, and does not prepare them well for schooling in later grades or for life generally.

Some problems seem to be common to many countries and most classrooms. Too much time is spent on rote learning, with the teacher talking or writing things on the blackboard and children repeating them or copying them down. Children generally do not work in groups on tasks that stimulate creative thinking or develop higher-order critical faculties. However, in many other respects, teachers—often, teachers in the same school, teaching comparable children—vary tremendously in their effectiveness.

Why are many teachers in Latin America and the Caribbean so much less effective than others? Is it mainly a problem of skills because practical tools that focus on teaching practices are not emphasized in pre-service and in-service training? Or is it mainly a problem of incentives because teachers in the region are generally not rewarded for better performance? These are important questions, as the likely effects of alternative policies depend on the reasons for the low performance of many teachers in the early grades in the region.

If teachers lack skills, professional development may help. Professional development can improve classroom practices and child learning outcomes. However, most professional development programs for teachers in the region are ineffective. They are theoretical rather than hands-on; they do not give teachers practical tools that help them become better teachers; and they are generic, rather than focused on a particular teacher's strengths and weaknesses.

Multiple studies from the United States show that teachers who receive coursework and personalized coaching on effective teaching practices can change their daily interactions with children in ways that have meaningful consequences for student learning and development.²⁴ The challenge is to find models for countries in the region that are effective, and can be taken to scale. This will require careful design and implementation, and rigorous evaluation, preferably

based on random assignment. With very few exceptions, this evidence has been lacking in the region.

It may also be, however, that some teachers do not exert enough effort. Under these circumstances, teacher pay for performance may help. Experimental and quasi-experimental evidence from the United States presents a mixed picture about the effectiveness of teacher pay-for-performance.²⁵ In developing countries, results have been more positive. Merit pay programs significantly improved learning outcomes in India (Muralidharan and Sundararaman 2011), Israel (Lavy 2002, 2009), and Mexico (Behrman and others 2015), although this evidence is largely about somewhat older children.

A number of countries and cities in the region, including Brazil (with Pernambuco) and Chile (with the Sistema Nacional de Evaluación del Desempeño, SNED), give rewards to teachers or schools (or both) that produce particularly large learning gains. However, in all countries in Latin America and the Caribbean the vast majority of a teacher's pay is determined by the number of years in service, and her contractual status (whether tenured or not).

Economists are generally more enthusiastic about pay for performance than educators. One question is what it is that teachers should be rewarded for. Most pay for performance schemes reward teachers on the basis of calculations of value added. Value added is a measure of the average increase in learning that takes place among plausibly comparable students assigned to different teachers (see the discussion in Box 5.2).

Box 5.2 Teacher Value Added

Estimating a teacher's value added is one way of measuring her effectiveness. Value added focuses on the learning gains among students taught by a teacher in a given grade. For example, to compare the value added of three teachers teaching first grade in different classrooms in the same school, one would estimate the average increase in child development or test scores between the end of kindergarten and the end of first grade, separately for children in each classroom. This is an estimate of the value added of each teacher. To see how much better one teacher is than another, one would also calculate the mean increase for all first graders

in that school. The difference between the learning gains of children in one teacher's classroom and the average learning gains across all three first grade teachers in that school would be an estimate of her relative effectiveness.

The estimation of teacher value added has been popular among economists since the pioneering work by Hanushek (1971) and Murnane (1975). In part, this work, and much that followed, was a response to a consensus among researchers that, although there were large differences in the effectiveness of teachers, the observed characteristics of teachers, including their experience, contractual status, and credentials, explained very little of these differences. By focusing directly on child learning outcomes, measures of value added sidestep the focus on teacher characteristics. This is both the strength and the limitation of this measure. If the assumptions that are necessary for estimates of value added to have a causal interpretation hold, then value added focuses directly on what really matters: child development or learning outcomes. But the assumptions may not always hold, and estimates of a teacher's value added are silent about what it is that one teacher does that makes her more effective than another.

The most important assumption in estimating value added is related to the unobserved characteristics of children. Children are not generally assigned to classrooms at random. Headmasters know which teachers are more effective than others. A headmaster who seeks to equalize outcomes within a school might assign the most difficult children to the best teachers. If the information that the headmaster uses to make these assignments is not adequately "controlled for," the true value added of the best teacher would be underestimated (because she received the most difficult students). Conversely, better teachers may have more bargaining power, and a headmaster who would like to retain the best teachers in a school might give those teachers easier children. In this case, the true value added of the best teacher might be overestimated (because she received easier students). Parents—particularly parents who are most aware of and interested in their children's learning—may also exert pressure to have their children in one or another classroom. But these parents are likely to have unobserved characteristics that themselves have an impact on learning gains, regardless of the teacher their children are assigned to. This too would introduce biases.

Because all the relevant characteristics of teachers and students cannot be measured, estimates of value added make one critical

assumption: that whatever may be the differences in the children assigned to one or the other teacher, they are captured by the “baseline” measure of learning or development (which, generally, is learning outcomes at the end of the previous grade). Put differently, estimates of value added assume that any difference in learning *gains*, as opposed to learning *levels*, can be attributed to teachers—especially when the comparison is limited to children taught by different teachers within the same school.

A great deal of research has gone into testing this hypothesis. An influential paper by Rothstein (2010) used rich data from North Carolina to show that *future* teachers predict *earlier* learning gains. This is an indication that teachers and children were not matched with each other at random, and that the characteristics of teachers and children that determined the match may bias estimates of value added. More recent work by Chetty, Friedman, and Rockoff (2014), however, argues that estimates of teacher value added are a good indication of that teacher’s effect on learning, uncontaminated by possible differences across students. For example, in one set of estimates, the authors focus on teachers who change schools. Headmasters in one school may have different objectives than those in another school, and the student population may vary a great deal between schools. Even so, when a teacher with a positive value added changes schools, the value added in the school she leaves goes down, and the value added in the school she joins goes up, by proportionately the same amount, on average. This suggests that a teacher carries her “value added” with her, and this is largely independent of the school, headmaster, or student body.

Another concern with estimates of value added is that, even if they are causal, these estimates appear to vary a great deal for the same teacher from one year to the next. Part of that variation may be because some teachers are more effective at teaching a particular group of students, and the composition of their classroom varies from one year to the next—even if this is by chance alone. It may also be that teachers have a particularly good or bad year. Finally, measurement error of various sorts will tend to dampen the correlation of estimated value added for a given teacher in different years. Put differently, the same teacher may be equally effective in one year and the next, but measurement error will make it appear as if her effectiveness has changed across years.

Rewarding teachers on the basis of value added is attractive because it focuses on what matters—learning—rather than on what does not matter—observable teacher characteristics. Pay for performance could also have other advantages. In Latin America, teachers do not appear to be underpaid overall, relative to other white-collar occupations (such as office workers). However, the distribution of teacher wages is compressed at the top, relative to that of workers in similar occupations, suggesting that the most effective teachers are getting paid too little to keep them in the profession (Mizala and Nopo 2012). A steeper salary scale, with pay increases and promotions depending in part on teacher performance, may help to bring more talented people into the teaching profession.²⁶

Nevertheless, rewarding teachers on the basis of calculations of value added is not simple. Some concerns are practical. Calculating value added is data intensive. If, for example, teachers in first through third grade are to be rewarded on the basis of value added, it would be necessary to apply tests at the end of kindergarten, first, second, and third grade to all students every year.²⁷ Moreover, value added is a noisy measure of teacher quality, as can be seen by the fact that estimates of value added for the same teacher can vary considerably from one year to the next (Araujo and others 2014 for Ecuador).

Other concerns are related to possible behavioral responses by teachers to the introduction of high-stakes testing. Pay for performance on the basis of value added could encourage teachers to cheat, teach to the test (rather than emphasizing learning more broadly), or focus on particular groups of students (e.g., those who are just below a given proficiency cutoff). Some of these issues can be mitigated with a careful design of the details of the pay for performance scheme (Neal 2011). Alternatively, it would be possible to reward teachers for classroom behaviors that predict learning, rather than test scores, although this approach also has important limitations.²⁸

Estimates of teacher value added could also be used to identify teachers who, year after year, produce very little learning and development among the children in their classrooms. If, after receiving high-quality in-service training, these teachers continue to

underperform, serious consideration should be given to dismissal or early retirement.²⁹ Dismissing teachers is controversial and politically difficult, but the stakes in terms of possible improvements in child learning and subsequent outcomes are very high. Reasonable estimates for Ecuador, which build on calculations for the United States (Hanushek 2009, 2011; Kane and Staiger 2002b), suggest that replacing the lowest-performing 10 percent of kindergarten teachers with average teachers would raise the wages of all affected cohorts by roughly 1.6 percent (see Box 5.3).

Box 5.3 How Much Does Teacher Effectiveness Matter?

Research from the United States estimates that replacing a low-performing teacher (a teacher at the 10th percentile of the quality distribution) with an average teacher would boost the lifetime income of each child in that class by approximately \$40,000, which, for a class of 25 children, is equivalent to an increase in total earnings of close to \$1 million (Hanushek [2009, 2011] and Kane and Staiger [2002a] provide estimates of comparable magnitude).

Are these results relevant for Latin America and the Caribbean? Estimates like these always involve making a number of assumptions, but some simple calculations suggest that the value of improving the effectiveness of the lowest-performing teachers (or replacing them with other teachers) in the region may also be substantial. Holding years of schooling constant, a 1 standard deviation increase in literacy skills in Chile has been estimated to increase average wages by 15 percent (Hanushek and Zhang 2006). In Ecuador, simulations suggest that replacing the lowest-performing 10 percent of kindergarten teachers with average teachers would increase mean learning outcomes in kindergarten by 0.11 standard deviations (Araujo and others 2014).³⁰ If the increase in learning carries over from kindergarten to adulthood, and if the estimates from Chile can be used to approximate the labor market returns in Ecuador, then replacing the lowest-performing 10 percent of kindergarten teachers with average teachers would result in an increase in wages of all affected cohorts by roughly 1.6 percent per year. Moreover, if the increase in teacher quality motivates children to stay in school longer, as seems plausible, then there would be an additional benefit because children acquire more schooling.

Improving quality is more difficult than increasing access. There is no one-size-fits-all policy for all countries. Nevertheless, a judicious combination of monetary incentives for outstanding teacher performance; innovative programs of in-service training, coaching, and mentoring; and dismissal for teachers who are persistently low performers holds promise in many settings.



Except where otherwise noted, this work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 3.0 IGO License. To view a copy of this license, visit <https://creativecommons.org/licenses/by-nc-nd/3.0/igo/>