

## Evaluating multigrade school reform in Latin America

Patrick J. McEwan\*

*Department of Economics, Wellesley College, Wellesley, MA, USA*

This paper describes three multigrade school reforms in Latin America: (1) Colombia's *Escuela Nueva*, (2) Guatemala's *Nueva Escuela Unitaria*, and (3) Chile's MECE-Rural. Each reform endowed primary teachers and students with special training and instructional materials, and encouraged new kinds of instruction in rural classrooms, with the goal of improving student learning. The paper discusses the challenges to evaluating school reforms in a rural context. It then reviews the international literature on each reform, assessing whether reforms were effectively implemented and whether they caused improvement in student outcomes. Even in the best cases, multigrade schools were not fully implemented. There is consistent evidence of positive effects on student achievement. Yet, the causal interpretation of evaluation findings is severely limited, especially in Guatemala and Chile. It concludes with suggestions for improving multigrade evaluations.

### Introduction

The Millennium Development Goals specify that children in every country should be able to complete a full course of primary schooling by 2015 (Birdsall et al. 2005). The Education for All initiative promulgates a similarly ambitious goal, with an emphasis on children in difficult circumstances, ethnic minorities, and girls (World Education Forum 2000; Little 2001). In Latin America, poor and indigenous children live disproportionately in rural areas (Winkler 2000; McEwan 2004). Yet, despite national progress in attaining goals, rural and urban children rarely obtain the same quantity and quality of formal schooling.

Figure 1 describes the difference in the net enrolment rates of urban and rural children (ages 6 to 18) across Latin America. Urban rates are 5 to 20 percentage points higher, especially in poorer Central American countries. Rural children also obtain lower average scores on international tests of academic achievement (even excluding the disproportionate number of rural children that do not attend school). A 1997 UNESCO assessment of students in early primary grades found that urban students score, on average, at least 20% of a standard deviation higher on tests of mathematics and language (Figure 2).

The archetypal rural school is located in an isolated area of low population density. It enrolls a small number of students with a range of ages and grades. The school is often staffed by no more than a single teacher in cramped infrastructure, and may not offer a complete primary education. Teachers usually receive no special training or materials to manage instructional time effectively in such classrooms, and instead emphasise large-group lectures and rote memorisation. This paper describes several varieties of multigrade school reform, a common strategy for improving rural primary schools in Latin America.<sup>1</sup>

---

\*Email: pmcewan@wellesley.edu

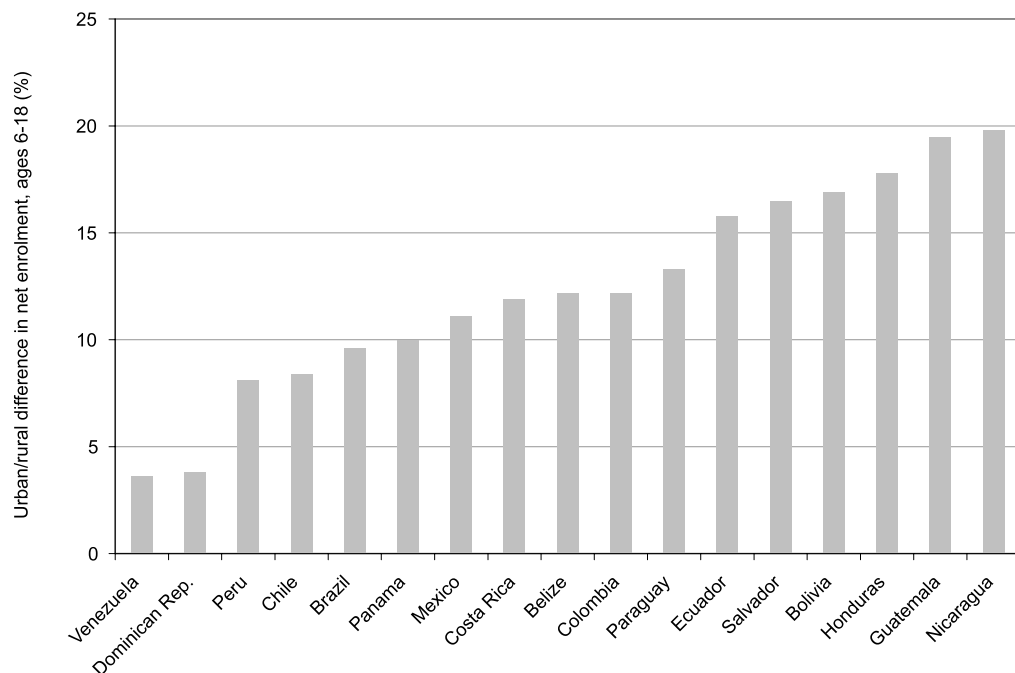


Figure 1. Urban-rural differences in net enrolment.

Source: Urquiola and Calderón (2006) and author's calculations.

Note: Each column indicates the difference between the net enrolment rate of urban children and rural children (ages 6–18) in any level of schooling.

The earliest multigrade reform in Latin America was Colombia's *Escuela Nueva*, though it has since been vigorously pursued in Guatemala and Chile (Colbert et al. 1993; Craig et al. 1998; García-Huidobro 2000). The reforms have been adapted differently, but they share several characteristics. First, they emphasise that teachers should receive adequate training to teach in a multigrade setting. Second, the reforms provide teachers and students with instructional materials and textbooks that are suited to a 'one-room schoolhouse' in which teachers cannot instruct all students and grades at the same time. Third, the reforms address the relative isolation of teachers by providing opportunities for professional interaction with colleagues in other schools. Fourth, the reforms emphasise that students are active participants in the learning process. As such, students are encouraged to work independently and creatively towards specified learning goals.

Multigrade reform is prominent in the educational recommendations of international organisations (PREAL 2003; UNESCO 2004; World Bank 2005). These conclusions are often based upon positive research findings, to be cited below, from reforms in Colombia, Guatemala, and Chile. The Latin American research asserts that reforms caused large and sustained improvements children's academic achievement and other outcomes such as democratic attitudes. These conclusions are usually adopted uncritically, despite shortcomings in the ability of non-experimental research to derive valid inferences about the causal impact of reforms. The goal of this paper is to describe and assess research conclusions from three countries, provide important caveats when necessary, and outline considerations for future evaluations of multigrade school reform. It bears emphasis that the paper only discusses Latin American reforms. Little (2006) reports detailed case studies and evaluations of multigrade reforms in several African and Asian countries.

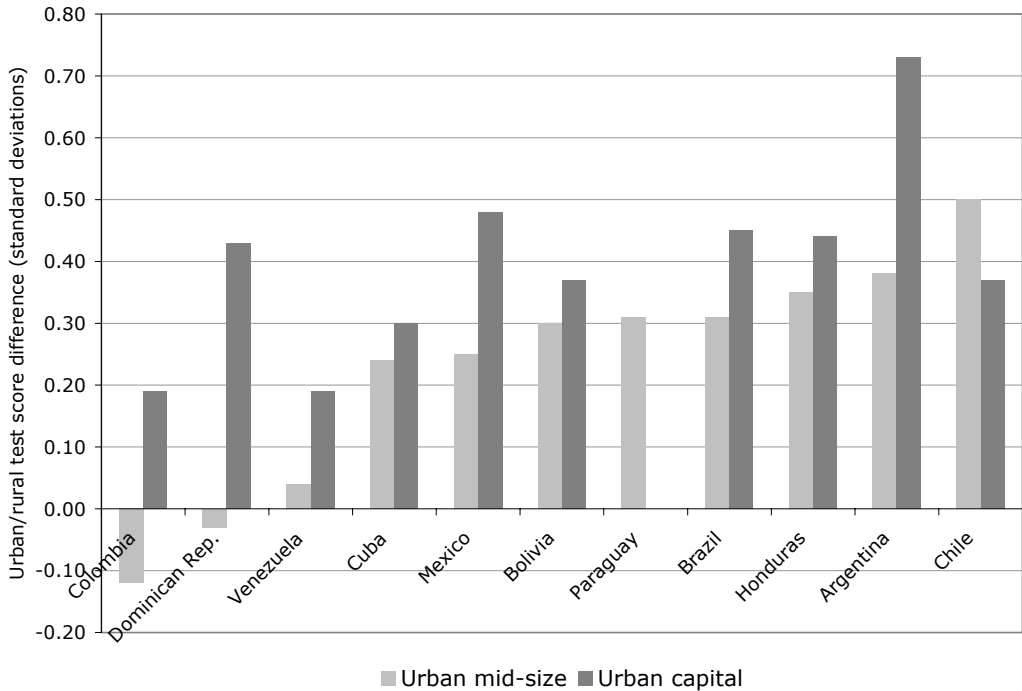


Figure 2. Urban-rural difference in test scores.

Source: LLECE (1998) and author's calculations.

Note: To calculate the gaps, I estimated the mean differences in achievement between urban students (in either mid-size or capital cities) and rural students, for each combination of grades (3 and 4) and subject (maths and language). The gaps reported for each country are the mean of the four gaps.

The paper proceeds in four steps. First, it defines 'multigrade' schooling and describes the features of three reforms in Colombia, Guatemala, and Chile. Second, it discusses challenges to evaluating school reform, some unique to rural contexts. Third, it summarises and critiques the research that informs two policy questions. First, did the reforms deliver the required training and materials, and did they change the classroom behaviour of teachers and students in desirable ways? Second, did the reforms produce sustained improvements in student outcomes such as academic achievement? The final section concludes and charts directions for future research.

### Background on multigrade school reform

The term 'multigrade' encompasses two separate but related characteristics of schools and their students. First, multigrade schools enrol students of multiple ages and grade levels in the same classroom. Second, these schools are the occasional recipients of an additional 'package' of training and instructional materials that are expressly designed to improve learning in heterogeneous classrooms.

#### *Age-grade heterogeneity in classrooms*

Multigrade schools are usually small, with as little as one classroom that is staffed by one teacher. They enrol children of multiple ages and grade levels in that classroom. In Latin

America, such schools are the usually born of necessity, given their location in sparsely populated rural areas.<sup>2</sup> Nonetheless, policy-makers can control the number and exact location of schools, and the degree of age-grade heterogeneity. A policy-maker could choose to construct a single rural school with 6 primary teachers in 6 grade-specific classrooms, each enrolling 10 students. Alternatively, the policy-maker could construct six separate and far-flung rural schools, each with a single classroom enrolling 10 local students that are heterogeneous in age and grade.

The decision entails trade-offs in access to primary schooling. By necessity, the larger school is more distant from some students. In rural areas, there are monetary and non-monetary costs to attending primary school, including the price of transportation; the time and energy used in walking to and from school; and the threats to personal safety that confront young children on such journeys.<sup>3</sup> If families in a large school's attendance area decide that costs are prohibitive, then the likely results are delayed enrolment, poor attendance, grade repetition, and drop-out.

There are also potential trade-offs in the quality of primary schooling. Age-grade heterogeneity has uncertain implications for student learning. By some accounts, older and more able students can serve as role models and tutors for younger children, developing their own skills in the process. This positive peer effect could be supplanted by a negative effect if older students intimidate or bully younger students, or if a particular group of students receives a disproportionate amount of teacher attention, either individually or in large-group instruction. The multigrade reforms described in the next section explicitly attempt to promote positive peer interactions via student collaboration in small groups, and to ensure that students work effectively even when they do not receive direct instruction from teachers.

The available data suggest that Ministries of Education in Latin America opt for numerous and small multigrade schools, instead of consolidated and distant ones. In Chile, 8727 schools offer primary instruction (grades one to eight). Of these, 53% are rural, though they enrol only 17% of primary students.<sup>4</sup> Table 1 confirms that rural schools have fewer classrooms that often enrol students from multiple grade levels. Eighty-six percent of urban schools have eight or more classrooms, and none of these classrooms is multigrade. In contrast, 47% of rural schools have only one classroom, of which 100% are multigrade. Another 22% of rural schools have two or three classrooms, and over 90% are multigrade.

In many countries, age-grade heterogeneity in primary schools is not complemented by special instructional approaches. Indeed, the norm is that rural schools receive the same instructional materials, and rural teachers receive no pre-service or in-service training in multigrade teaching. Because of this, many observers refer to the 'invisibility' of multigrade schools (Little 2001; Williamson 2004). Since the 1970s, several Latin American countries have implemented wide-ranging interventions that directly improve the quality of rural, multigrade schools.

### ***Multigrade reforms in Colombia, Guatemala, and Chile***

Colombia's *Escuela Nueva* is the pioneering reform, and often cited as a best-practice in rural school reform (Schiefelbein 1991; PREAL 2003). Its roots are in the unitary schools movement of the 1960s, but it was first implemented on a large scale in 1975, with support from the US Agency for International Development (USAID).<sup>5</sup> During the next decade, it continued to expand with financing from the Coffee Growers' Federation, the Inter-American Development Bank, and other organisations. In 1986, with support from the World Bank, the intervention was codified in a 'kit', in order to support its large-scale

Table 1. Urban and rural schools in Chile, 2000.

Number of classrooms	Urban primary schools (%)	Rural primary schools (%)	Average % of classrooms in each school that are multigrade	
			Urban	Rural
1	1.0%	46.5%	40%	100%
2	1.4%	15.4%	20%	99%
3	1.4%	6.1%	36%	93%
4	2.5%	4.1%	12%	51%
5	2.3%	6.1%	17%	57%
6	3.1%	2.7%	4%	27%
7	2.0%	1.8%	2%	13%
8 or more	86.4%	17.4%	0%	0%
Total %	100%	100%		
Total number	4,087	4,640		

Source: Ministry of Education enrolment file and author's calculations.

Note: The data include all schools, public and private, that provide instruction in grades 1 to 8.

expansion. It has received ongoing support from the private sector, often through official letters of agreements with the Ministry of Education (Perfetti 2004).<sup>6</sup>

In Guatemala, the *Nueva Escuela Unitaria* is a more recent innovation, borrowing ideas and personnel from the Colombian reform. In 1989, the Ministry of Education and USAID started Project BEST to improve the quality of primary schooling. As one component, it implemented a multigrade school intervention in 1993. The pilot programme focused on 100 schools in several regions, expanding to another 100 schools in the next two years. In later years, additional school expansions received financial support from non-governmental organisations such as the coffee growers of Guatemala (Craig et al. 1998; Juárez and Associates 2003).

Finally, Chile implemented its own rural school intervention, MECE-Rural, as part of a nationwide primary school reform in the 1990s that emphasised the reduction of inequality (García-Huidobro 2000; Cox 2004).<sup>7</sup> Until 1992, multigrade primary schools and students received no special attention from the Ministry of Education. Beginning in 1992, the Ministry provided training and materials to rural schools, reaching all multigrade schools by 1996. By the end of the decade, MECE-Rural was fully incorporated into Ministry operations and rechristened the Rural Basic Education Programme (Ministerio de Educación, n.d.).

The general features of the Colombian, Guatemalan, and Chilean reforms are described in Table 2. Each reform has much in common, particularly the Colombian and Guatemalan versions. First, each intervention emphasised in-service training for current rural teachers, rather than pre-service education. Teachers learned instructional techniques suitable for a multigrade setting, including individual and cooperative learning, as well as the use of newly-developed instructional materials. The training itself applied the instructional techniques, providing a model for eventual use in classrooms. Much of the training was conducted within local groups of rural teachers, referred to as 'microcentres' or 'teacher circles' depending on the context. These provided a venue for collegial exchanges and ongoing professional development.

Second, each intervention promoted the development and distribution of multigrade instructional materials, including teacher guides and student textbooks that facilitated

Table 2. Multigrade school reform in Colombia, Guatemala, and Chile.

	Colombia: <i>Escuela Nueva</i>	Guatemala: <i>Nueva Escuela Unitaria</i>	Chile: MECE-Rural
Training and professional development	Multiple in-service training workshops (initiation, use and adaptation of textbooks, use of library, follow-up); demonstration schools; micro-centres to support local rural teachers	In-service training workshops in instructional methods; 'Teachers' Circles' meetings for professional development	In-service training workshops provided through 'micro-centres' of local rural teachers; supervisors provide instructional support
Materials	Teacher guides; student textbooks that guide independent work; classroom libraries; learning activity centres in classrooms	Teacher guides; student textbooks that guide independent work; classroom libraries	Revised teacher manual for multigrade rural and multigrade setting
Instruction and pedagogy	Independent and small-group activities; peer collaboration; learning activities inside and outside the classroom; little use of large-group instruction	Independent and small-group activities; peer collaboration; learning activities inside and outside the classroom; little use of large-group instruction	Independent and small-group activities; active learning using spaces inside and outside of classroom; 'Educational Improvement Projects' based in micro-centres
Student involvement	Student government and committees guide school functions; emphasis on democratic participation	Student government and committees guide school functions; emphasis on democratic participation	
Community involvement	School-based gathering of family and community information; agricultural calendars	Community map; agricultural calendars; oral histories	
Student evaluation	Continuous, student-specific assessment of learning; students advance through units at individual pace ('flexible promotion')	Continuous, student-specific assessment of learning; students advance through units at individual pace ('flexible promotion')	
References	Schiefelbein (1991); Colbert et al. (1993); McEwan (1998)	Craig et al. (1998); Kraft (1998)	Ministerio de Educación (n.d.); García-Huidobro (2000); Cox (2004); PREAL (2003); Williamson (2004)

self-guided learning. The texts proceed in units that students can pursue at their own pace and without continual teacher supervision. Third, the training and materials emphasised the application of active pedagogies. Students participate in individual and small-group exercises that require the acquisition and application of new skills, rather than passive attendance in a large-group lecture. The exercises are often relevant to the rural and agricultural settings of the communities, and may involve activities inside and outside the classroom.

Fourth, governments and committees involve students in the management and organisation of the school, particularly in the Colombian and Guatemalan versions. The broad goal is to instil students with cooperative and democratic attitudes, consistent with the pedagogical approach. Fifth, the interventions in Colombia and Guatemala provide specific mechanisms for involving the community in the school, including the development of local agricultural calendars. Sixth, student evaluation occurs throughout the year, emphasising mastery of each unit's skills, in place of a traditional 'pass or fail' assessment at the year-end. Because of this, students do not necessarily need to repeat the entire school year if they are absent for some days, an important consideration in agricultural areas. Colombian and Guatemalan sources describe 'flexible promotion' as a key strength of the intervention, given higher rates of grade repetition and drop-out in rural areas (Colbert et al. 1993; Kraft 1998).

### Issues in the evaluation of multigrade schools

Before describing evaluations of multigrade reforms, I review three challenges inherent to impact evaluations of rural school reform in Latin America: (1) the adequate definition of the multigrade 'treatment' to which students are exposed; (2) internal validity, or whether a study credibly identifies a causal relationship between the treatment and student outcomes; and (3) external validity, or whether a study's findings can be generalised to rural schools and students not included in the evaluation.<sup>8</sup>

#### *Defining the treatment*

The *sine qua non* of an evaluation study is a clear description of the 'treatment', or the intervention that is being evaluated. In the absence of such a description, it is problematic to interpret the study's results, or to generalise its findings to other schools. Table 2 provided a general description of the ideal multigrade school treatment in three countries, including training, materials, and pedagogical approaches, with a particular focus on the 'package' of educational inputs. In several regards, this description is incomplete.

First, multigrade school interventions often occur in the context of larger systemic education reform. It is important to describe whether schools received additional treatments, beyond a discrete multigrade intervention. In Chile, for example, MECE-Rural was just one component of a much larger reform that emphasised a revised nationwide curriculum, improved textbooks and school infrastructure, and added computers and networking (Cox 2004). A subcomponent of the reform – the P-900 programme – targeted low-achieving schools, some of them in rural areas, with remedial tutoring and teacher training (García-Huidobro 2000; Chay et al. 2005). More recently, the Ministry of Education has promoted bilingual education in predominantly indigenous schools in the southern regions, many of which belong to MECE-Rural. The initial phase of both P-900 and MECE-Rural in the 1990s ignored linguistic and cultural diversity among students, unless specifically adapted by local participants (Herrera Lara 1999).

A similar phenomenon is evident in Guatemala. In the early stages of implementation, the pilot of *Nueva Escuela Unitaria* was implemented in 100 Ministry of Education schools. In subsequent phases, schools were operated by Catholic groups, international non-governmental organisations, and the private coffee grower's association (Craig et al. 1998). In other instances, multigrade schools are operated by associations of parents and community members via Guatemala's PRONADE programme that provided legal and financial mechanisms for the steady expansion of rural schooling (McEwan and Trowbridge 2007). In each case, the multigrade treatment may differ depending on how rural schools are managed and financed. The early evaluations of the Guatemalan programme are limited to pilot schools operated by the Ministry of Education, avoiding this dilemma.

Second, it is almost never accurate to speak of a single, homogeneous treatment, except in a highly-controlled, small-scale experiment. Rather, there is wide variation in how treatments are actually implemented in schools, a common feature of education policies (McLaughlin 1991; Loeb and McEwan 2006). Training is not always provided (or falls short of its goals), materials do not arrive in classrooms, and entrenched pedagogies are resistant to change. Even in nominally successful interventions like the Colombian *Escuela Nueva*, many teachers' practices resemble those of traditional rural schools, and some 'traditional' teachers independently adopt multigrade practices (Benveniste and McEwan 2000; Forero-Pineda et al. 2006). Understanding variation in implementation helps to explain the intervention's effects on student outcomes, or lack thereof. It also provides information that may assist in the subsequent expansion or modification of the intervention.

Third, the description of the treatment entails more than an inventory of tangible resources like textbooks and training workshops. The initial stages of the Colombian and Guatemalan reforms emphasised the direct involvement of teachers in the design and modification of the programme. McGinn (1996) observed that the process itself created a sense of 'ownership' among teachers, many of them highly-motivated volunteers, and was a driving force behind its early success. This aspect of the treatment is far more difficult to describe and replicate than a series of discrete inputs. This may explain why small-scale success has proven challenging to replicate on a large scale, especially among non-volunteers (Benveniste and McEwan 2000; McEwan 2001).

Fourth, the initial treatment could include Hawthorne effects, in which rural schools respond to the mere presence of outside collaborators and researchers. In any country, rural schools and their teachers receive less attention than most. The arrival of Ministry of Education personnel, along with good-will or merely added pressure, is itself a form of treatment. Thus, it is important to follow rural schools beyond the initial years of the treatment and probe for evidence of Hawthorne effects that are neither sustainable nor representative of large-scale effects. In the case of Colombia, the evaluations were conducted after the programme had existed for many years, making Hawthorne effects less plausible (Psacharopoulos et al. 1993; McEwan 1998). In contrast, the Guatemalan evaluations focused on a small number of schools in the early stages of the programme (de Baessa et al. n.d.).

### **Internal validity**

A key goal of multigrade impact evaluations is to determine whether exposure to the intervention *causes* increases in student outcomes. If it does, then the evaluation is said to possess internal validity. To make this determination, researchers would like to observe a single group of students in two states of the world: (1) a state in which students are exposed to the intervention, and (2) a state in which they are not. It is physically impossible to



observe both at once. Instead, researchers compare the outcomes of students exposed to an intervention (the treatment group) with those of a different group of students who are not (the control group). This comparison provides an internally valid estimate of the causal effect of the intervention, *as long as students are similar in all regards except for their exposure to the intervention*. If students are different, then the evaluation suffers from a common malady known as selection bias. In such a case, it is difficult to determine whether the intervention is truly responsible for group differences in student outcomes.

The best way of ensuring the equivalence of treatment and control groups is to randomly assign the treatment to students or schools (Shadish et al. 2002). There are a growing number of attempts to conduct randomised, experimental evaluations of education reforms in developing countries (Glewwe and Kremer 2006). However, no experiments have been conducted on multigrade school interventions. As an alternative, researchers have compared students attending treated schools to those in 'similar' schools, as defined by their location, age-grade heterogeneity, poverty, or other characteristics that are not always described in evaluation reports. Sometimes the researchers use statistical techniques such as regression analysis to control for some observed attributes of students in the two groups (Psacharopoulos et al. 1993; McEwan 1998). Other researchers compare changes in school-level or student-level test scores over time, in order to control for level differences in student outcomes that are not caused by the treatment (Universidad Austral and Universidad de Playa Ancha 1998).

Non-experimental studies are susceptible to several varieties of selection bias, resulting from unobserved, pre-intervention differences between the treatment and control groups. These differences might affect the level or growth of student outcomes, independently of the multigrade school interventions.

First, governments and NGOs are typically responsible for deciding which schools receive interventions.<sup>9</sup> The administrative logic is rarely stated, and could bias estimated impacts in positive or negative directions. On the one hand, multigrade schools are allocated to isolated schools with a small number of teachers. These schools are usually located in poor areas in which students have low outcomes. If control schools are drawn from less isolated and less impoverished areas, then the impact of the multigrade intervention on student outcomes will be understated. Poverty drags down the mean outcomes of treated students, but not students in the control group.

On the other hand, the early implementation of multigrade interventions in Colombia and Guatemala was targeted at volunteer schools. Unless control schools are drawn from a similar group of volunteers – and the extant literature offers no evidence of this – then it introduces a positive bias. An apparently successful intervention may be due to the pre-existing motivation of an outstanding group of teachers and administrators, rather than the intervention per se.

A second source of selection bias is related to the characteristics of students that attend treatment and control schools. Putting aside the previous discussion, let us suppose that students in the two school types are similar at the beginning of the intervention (i.e., treatment and control schools are chosen randomly from the population). There are two ways in which the treatment could introduce differences between students, and therefore in their measured outcomes at the end of the school year.

First, it is probable that students, after the evaluation has begun, are more likely to choose better-equipped multigrade schools. If highly-motivated students also have better outcomes, then it would artificially inflate the measured outcomes of treatment schools. Nonetheless, parental choice between rural schools is not a strong possibility in most contexts. Even when students are permitted to choose among public schools, as in Chile, there is usually only one rural school within a reasonable distance of their home. Some

authors have noted that this guards against student selection bias (Psacharopoulos et al. 1993; McEwan 1998).

However, there is a second possibility in systems that cannot boast of universal primary enrolments such as Guatemala (McEwan and Trowbridge 2007). Suppose that 10% of rural students do not attend their local rural school, whether it belongs to the treatment or control group. Presuming that the multigrade intervention improves quality in treated schools, it may induce non-attending students to enrol.<sup>10</sup> Indeed, that is an explicit goal of multigrade interventions. Yet, the newly-enrolled students – if more impoverished or less ready for school – will lower the average achievement in the treatment group relative to the control group. Thus, the impact of the multigrade reform on academic achievement is biased downward. In a similar vein, an effective multigrade school could prevent drop-outs among students on the margin of doing so. This could pull down average student achievement in upper primary grades, a possibility suggested by Psacharopoulos et al. (1993) and McEwan (1998).

### *External validity*

Impact evaluations are intended to inform policy, often preceding a decision about whether to ‘scale up’ the intervention to all rural schools. If the evaluation results can be easily generalised to other schools, then it is said to possess external validity. Yet, there are obvious limits to external validity, even if results are drawn from internally valid evaluations.

As noted above, schools and teachers that participate in pilot interventions are volunteers. The essential question is whether a positive impact among 10 or 20 volunteer schools can also be achieved among 1000 schools, some of which are indifferent or even resistant to the multigrade intervention. In the best cases, programmes have explicitly acknowledged this challenge in the process of scaling up. For example, the Colombian *Escuela Nueva* emphasises visits by prospective school personnel to model schools, cultivating understanding and enthusiasm (Colbert et al. 1993). In the worst cases, materials and workshops are literally delivered to schools, with little effort to involve personnel in the planning or management of the intervention. It seems unlikely in such cases that evaluation results from small-scale pilot programmes are a valid indicator of large-scale impacts.

A second challenge is economic, cultural, and linguistic heterogeneity among rural students. A pilot evaluation could succeed in demonstrating positive effects among a subgroup of students. But others could respond differently to the multigrade treatment. A primary example is the substantial population of indigenous students in southern Chile and several regions of Guatemala. In the early implementation of MECE-Rural, there was little consideration of student ethnicity (Herrera Lara 1999). In some cases, programme developers allow for adaptations of training and textbooks to suit local conditions, often involving teachers in the process (Colbert et al. 1993). Also, evaluators have attempted to include diverse student populations in pilot evaluations (de Baessa et al. n.d.).

## **How well were reforms implemented?**

### *Colombia*

This section assesses whether ‘ideal’ reforms were actually implemented in schools, relying on descriptive evidence from evaluation studies. It focuses on two aspects of implementation: (1) Did resources such as training and textbooks, judged to be essential for implementation,

actually arrive in schools? (2) Did teachers modify their instructional strategies in response to the reform?

In one of the first large-scale evaluations of *Escuela Nueva*, Rojas and Castillo (1988) reported that the majority of treated schools applied the basic set of inputs.<sup>11</sup> Ninety-four per cent had a school library, 98% operated 'learning corners' within classrooms that complemented student exercises in the textbooks, and 97% had prepared a 'community map' to facilitate local school-community partnerships. In some respects, schools did not have a full complement of resources: 67% reported having sufficient self-instructional textbooks, 58% operated the student government, and 61% maintained an agricultural calendar. These figures were drawn from a sample of schools that had implemented the model for at least five years.

Using a more representative survey of treated schools from the early 1990s, McEwan (1998) reported that only two-thirds are using a school library. Fewer than half of the schools reported using the self-instructional textbooks, a key programme element. In a sample of 425 *Escuela Nueva* teachers, Benveniste and McEwan (2000) reported that about three-quarters had participated in the three training workshops that introduce the instructional methodology and its materials.

Given the absence of materials and training in some schools, it would be unsurprising if some teachers did not alter their teaching practices. Benveniste and McEwan (2000) tested this by comparing teacher-reported usage of instructional techniques in a sample of *Escuela Nueva* and traditional rural schools. They constructed an index of teaching methods used in *Escuela Nueva*: individual problem-solving by students, exploration of spaces outside the classroom, manipulation of objects (as in the use of learning corners), and the use of small student groups. *Escuela Nueva* teachers had index values that are one-third of a standard deviation higher than traditional teachers, a statistically significant difference. It is possible that treated schools attracted teachers that were predisposed to use such instructional techniques. However, the strongest predictor of the use of such instructional techniques within the sample of *Escuela Nueva* teachers was participation in the initial workshop on textbook use, suggesting that capacity-building plays an important role.

### Guatemala

Chesterfield and Rubio (1997) analysed 13 teachers in the pilot schools of *Nueva Escuela Unitaria*. They found that two-thirds of the sample, depending on the year, had access to materials like learning corners, libraries, and textbooks. By the fourth year of the programme, more than 80% of teachers had received training in maths and Spanish instruction. However, the small sample size casts doubt on the precision of these estimates. In a separate study, de Baessa et al. (2002) observed the frequency of teaching practices in a small sample of 10 programme schools and 10 non-participating rural schools. In general, students in the *Nueva Escuela Unitaria* spent a much larger proportion of the school day in small group settings, both with and without teacher support. Students in traditional rural schools spent the majority of time carrying out individual 'seatwork'.

### Chile

After the first five years of MECE-Rural, an official evaluation surveyed teachers regarding the availability of instructional materials (Universidad Austral and Universidad de Playa Ancha, 1998). Among participating teachers, 65% and 72% reported that textbooks and teacher guides, respectively, arrived in classrooms on time. The vast majority of participating teachers (94%) had participated in at least some training.

A representative sample of MECE-Rural participants confirmed these findings. Of the most isolated rural schools, about two-thirds received student textbooks during the 2001 school year (CIDE 2002). While most classrooms had received classroom library resources, 87% of teachers reported that libraries were incomplete. Fifty-four percent of teachers received training that specifically emphasised the use of new instructional materials like self-instructional textbooks. There was no evidence on whether Chilean rural teachers actually modified their teaching practices in response to the availability of training and materials.

## Did reforms affect student outcomes?

### Colombia

Several impact evaluations have been conducted on the *Escuela Nueva* multigrade intervention, using multiple sources of data (Rojas and Castillo 1988; McEwan 1998; Psacharopoulos et al. 1993). The initial evaluation, in the late 1980s, gathered data on third-grade and fifth-grade student outcomes in a sample of 168 schools that had participated in the intervention for at least five years (Rojas and Castillo 1988). These were matched to a quasi-experimental sample of 60 traditional rural schools in similar geographic areas. The sample was re-analysed, by Psacharopoulos et al. (1993), using more elaborate statistical methods. The evaluations show that students in the treatment and control groups are similar, though *Escuela Nueva* students have slightly *lower* levels of poverty, which might also affect outcomes. This highlights the importance of making statistical controls for observed differences across the groups.

Upon comparing student outcomes in a regression framework, Psacharopoulos et al. (1993) found effects of 0.31 to 0.37 standard deviations in third-grade Spanish and maths tests. The effects exceed recent estimates of other successful educational interventions, such as Chile's P-900 programme (Chay et al. 2005). The evaluation also compared students' civic behaviour, creativity, and self-esteem, and found that the intervention only had positive effects on civic behavior.

The results are subject to several caveats. The data are cross-sectional, gathered at a single point in time. In this regard, they are similar to many other studies of education quality in developing countries (Glewwe and Kremer 2006). In all such studies, it is likely that students in treatment and control schools differ in unobserved ways that cannot be controlled for statistically. Given the previous discussion, a particular concern is that the *Escuela Nueva* prevents student drop-outs in upper grades, or induces some students to re-enter primary school (Psacharopoulos et al. 1993), perhaps biasing downward the measured programme impact.

A related concern is that the initial evaluation was based upon schools in which the programme was relatively well implemented, perhaps including the most enthusiastic supporters of the intervention. This evokes concerns about internal validity, since the best-implemented, volunteer schools are likely to perform better than traditional rural schools even in the absence of the treatment, biasing upward the results. Even if the results are internally valid, it evokes concerns about external validity, since the findings from volunteer schools may not be generalisable to other schools.

Fortunately, the results of a second evaluation can be used to corroborate and extend the initial findings. McEwan (1998) analysed a representative sample of all rural schools in the Pacific region of Colombia, collected in 1992 by the official Ministry of Education assessment system.<sup>12</sup> Controlling for a wide range of student and school characteristics, the analysis found that *Escuela Nueva* third-graders scored 0.26 to 0.45 standard deviations

higher in Spanish and maths, respectively. The effects were not statistically significant in the fifth-grade, consistent with the previous evaluation. It is subject to the same caveats regarding internal validity, but its results may have greater external validity (at least within Colombia's Pacific Region) because the sample is representative of all primary schools.

Lastly, a more recent evaluation compared proxy measures of peaceful student interaction across schools that implement the *Escuela Nueva* strategies in varying degrees (Forero-Pineda et al. 2006). In a sample of 25 primary schools in the coffee-growing region, the study found that an index of implementation was positively associated with a measure of peaceful interactions among children, even after controlling for other student and school variables in a multilevel model.

### Guatemala

The *Nueva Escuela Unitaria* (NEU) is the subject of several evaluations financed by USAID and conducted by private contractors. They are cited approvingly in descriptions of the treatment, but with no consideration of evaluation criteria such as internal and external validity (Craig et al. 1998; Kraft 1998). Unfortunately, the validity of the non-experimental evaluation results is difficult to assess, because the evaluations provide vague and sometimes contradictory information about the evaluation design, exact samples of data, and the statistical techniques used to analyse them.<sup>13</sup>

The achievement evaluation was conducted between 1993 and 1996 during the pilot stage of *Nueva Escuela Unitaria* (de Baessa n.d.; de Baessa et al. n.d.). The evaluation sample consisted of children in the early grades of ten treated schools and ten control schools, divided between two regions of Guatemala (one region included a large proportion of indigenous students). The control schools were apparently chosen for their similarity to treated schools (de Baessa et al. 2002), but there is no pre-test or baseline data that are reported to confirm this assertion. The impact findings are disaggregated by year and region for each achievement test, with no statistical controls for baseline differences in student characteristics. Given the already small number of schools, this leads to extremely small samples. There are many statistically insignificant comparisons, and a robust pattern of findings does not emerge.<sup>14</sup> The findings are subject to all of the critiques described previously. Selection bias is a particular concern given the small, self-selected sample of treated schools, the absence of student controls, and the cross-sectional comparisons. In short, the evaluations do not provide credible evidence on programme impacts.

A separate evaluation used Guatemala's official education assessment, the PRONERE, to compare achievement in NEU and traditional rural schools (Juárez and Associates 2003). The sample was larger and more representative. The analysis found that average achievement among separate cohorts of sixth-graders increased more rapidly between 1998 and 2001 in the sample of treated schools, compared with other rural schools. The study does not confirm whether these different trends could be explained by changing characteristics of students, despite the availability of such variables in the PRONERE data (McEwan and Trowbridge 2007). Moreover, the treated schools were volunteers (Craig et al. 1998), and may have experienced larger test score gains even in the absence of the treatment. The results are more convincing than the pilot evaluation, but fall short of providing credible evidence.

### Chile

There have been two efforts to assess the impact of MECE-Rural on the academic outcomes of rural students (Universidad Austral and Universidad de Playa Ancha, 1998;

García-Huidobro 2000). Both utilised fourth-grade test scores from Chile's National System for Education Quality Measurement (SIMCE), which covered the majority of primary schools. However, the smallest and most isolated schools, including many of the schools targeted by MECE-Rural, did not appear in the SIMCE data.

García-Huidobro (2000) estimated the growth in fourth-grade test scores that occurred in 196 MECE-Rural schools between 1992, the programme's first year, and 1996 (omitting any control group comparisons). However, it is problematic to extract any causal conclusions from simple 'before-and-after' comparisons (Shadish et al. 2002). First, the study did not employ a control group of rural schools that did not participate in the treatment. As a result, it is difficult to ascertain whether MECE-Rural schools would have experienced test score growth even in the absence of the treatment. One reason to suspect such growth is that the SIMCE tests in the early 1990s were not statistically equated. Thus, test score gains, a common finding for most schools during this period, could be an artefact of declining test difficulty. Another possibility is that test scores rose because of a general increase in educational expenditures, rather than a programme-specific effect. During the 1990s, the Ministry of Education substantially increased the per-student payments in all publicly-funded schools (Cox 2004).

Second, there were no attempts to make statistical controls for the changing characteristics of students attending these schools.<sup>15</sup> Third, the MECE-Rural schools in the sample were relatively larger ones (since the smallest ones did not participate in the SIMCE assessment). Thus, the effects are hard to generalise to the entire population of MECE-Rural schools, especially the poorest and most isolated ones.

The second evaluation is a minor improvement, because it compared achievement growth in MECE-Rural schools to that of a control group that included all urban schools (Universidad Austral and Universidad de Playa Ancha 1998). It showed that test score changes are larger among rural students, though without presenting tests of statistical significance or a discussion of the magnitudes of the effects. It is doubtful that the urban control group reasonably approximated to how rural students would have fared in the absence of MECE-Rural. First, urban students were themselves subject to a wide-ranging series of reforms during this period, which may have lifted their test scores, leading to an under-statement of MECE-Rural's impact (Cox 2004; Chay et al. 2005).

Second, the test score comparisons used inconsistent samples of urban and rural schools in each year, complicating the interpretation of gains. For example, the urban sample size increased over time, apparently because the coverage of the SIMCE assessment increases. Sample composition could affect test score gains in the control group, independently of programme effects. There was no acknowledgement of these analytical challenges, or any attempt to probe the sensitivity of results to alternate samples.

As the previous sections have emphasised, Latin American research has infrequently used longitudinal test score data, especially on individual students. However, multigrade research in other countries often provides better examples in this regard, including Sri Lanka (Vithanapathirana 2006), the Turks and Caicos Islands (Berry 2001, 2006), and Vietnam (Vu and Pridmore 2006). More generally, Little (2006) provides an extensive review of the literature on multigrade schools outside Latin America.

## Conclusions

This paper has reviewed evaluation findings on three multigrade school reforms in Latin America. The first reform, Colombia's *Escuela Nueva*, is one of the oldest, largest, and most-imitated. Guatemala's *Nueva Escuela Unitaria* was directly inspired by the Colombian

reform, though it has never been implemented on the same scale. The final reform, Chile's MECE-Rural, was the rural component of a larger, system-wide education reform aimed at improving educational equity.

A consistent finding in Latin American research is that participating schools do not fully implement the multigrade strategies. In the most severe instances, schools do not receive the prescribed instructional materials, and teachers have not received training. Participating teachers in Colombia and Guatemala appear to teach differently, on average, engaging in more small-group instruction and using other elements of active pedagogy. However, there is still considerable variation in teaching methods.

In Colombia, the variation is partly associated with the availability of training (Benveniste and McEwan 2000). However, the reasons for poor adoption of new teaching methods are complex and little-researched. It may, in part, stem from the implementation process: 'top-down' methods among non-volunteer teachers encounter greater resistance than the 'bottom-up' approach favoured by innovators in the initial stages of *Escuela Nueva* and *Nueva Escuela Unitaria* (McGinn 1996; Benveniste and McEwan 2000). It bears emphasis that such implementation dilemmas are not unique to multigrade school reforms (Loeb and McEwan 2006).

The most credible impact evaluations are from Colombia, and they show consistently positive effects on academic achievement in the early primary grades, but less so in the upper primary grades and for non-academic outcomes (Psacharopoulos et al. 1993; McEwan 1998). Another study finds some effect on the alternative outcome of peaceful social interaction among students (Forero-Pineda et al. 2006). The studies make controls for student characteristics, but school and student selection bias are still likely threats to internal validity. Given the large-scale implementation of *Escuela Nueva*, threats to external validity are less problematic. Overall, these results suggest that regional inequities might be overcome through concerted investments in multigrade schools.

The best evidence from Guatemala and Chile compares achievement gains across treatment and control schools, and shows that gains are larger in treated schools (Universidad Austral and Universidad de Playa Ancha 1998; Juárez and Associates 2003). These studies do not acknowledge or probe the considerable threats to the internal validity of this conclusion, including biases from school or student selection, rendering some conclusions suspect. The quality of their quasi-experiment and statistical analysis is disappointing, given that the Guatemalan evaluations were conducted by contractors to USAID with access to financial resources.

Many of these challenges could be surmounted in future programme replications and evaluations. In each country, the reforms were scaled up as new cohorts of schools entered the programme in successive waves. In general, the process of entry is governed by unstated, arbitrary, or changing criteria (especially volunteerism among personnel or ease of access). By exercising deliberate selection procedures – perhaps randomised assignment – researchers can diminish the likelihood of selection bias and produce results that are internally valid and more capable of informing policy decisions.

A frequent critique of such procedures is that they are inconsistent with programme goals. For example, the success of multigrade interventions may actually hinge upon the enthusiasm of volunteers. However, this can easily be reconciled with modern evaluation designs. In an evaluation of whole school reform in Chicago public schools, Cook et al. (2000) identified a group of schools in which staff had indicated a willingness to participate in a whole-school reform. Of this group, a number of schools were randomly selected to receive the treatment. This necessarily reduces external validity, since the impact can only be generalised to other volunteer schools. However, a carefully implemented

randomised experiment preserves internal validity, and yields credible estimates of programme effects.

Conducting such an evaluation would not necessarily involve greater costs than the elaborate evaluations already funded by USAID (e.g., de Baessa et al. n.d.). It would require that evaluations are designed before the programme itself is implemented, in close collaboration with programme personnel. Glewwe and Kremer (2006) report numerous examples of such randomised experiments in African and Asian countries.

### Acknowledgement

I am grateful to Carlos Rojas, Emiliana Vegas, and Eduardo Velez for helpful conversations. I received financial support for this research from the World Bank, but I am solely responsible for its conclusions.

### Notes

1. On multigrade school reform in other countries, see Little (2006) and the citations therein. Other common rural school interventions in Latin America include cash transfers to poor families that are conditioned on school attendance (Rawlings and Rubio 2005), and the rapid expansion of community-managed schools in isolated rural areas (Jimenez and Sawada 1999; McEwan and Trowbridge 2007).
2. It is possible to purposely mix students of various ages and grades in urban settings, though evidence suggests that this is rarely pursued in developing countries. Many urban schools, particularly in poor areas, are de facto 'multi-age' schools, given high levels of delayed enrolment and grade repetition, which lead to a wider age distribution within grade-specific classrooms.
3. These threats are considerable in many rural areas of some countries, especially Colombia and Guatemala, with histories of rural violence and civil war.
4. Calculations are based upon the Ministry of Education's administrative enrolment file in 2000.
5. For details on the history of the *Escuela Nueva* reform, see Colbert et al. (1993), McEwan (2001), and the citations therein.
6. One of these organisations, the *Fundación Escuela Nueva Volvamos a la Gente*, is directed by a programme founder, Vicky Colbert.
7. Many low-achieving schools also participated in the P-900 programme, consisting of remedial tutoring and teacher training (García-Huidobro 2000; Chay et al. 2005).
8. For general discussions of these issues in evaluation, see Shadish et al. (2002).
9. Economists refer to this as endogenous programme placement (Pitt et al. 1993; Glewwe and Kremer 2006).
10. Equivalently, it may improve student attendance, and increase the likelihood that some students are in class on the day that an achievement test is administered to students.
11. Also see McEwan (1998), who cites data from the original evaluation.
12. A separate evaluation analyses a 1993 sample of rural students, also drawn from the official Ministry of Education assessment (Misión Social 1997). Its results are apparently consistent with other evaluations, but the exact statistical specifications and controls are hard to discern from the reported results.
13. For example, the evaluations refer to the use of a longitudinal design, even though achievement comparisons seem to occur within cross-sections and do not employ the panel structure of the data (de Baessa n.d.; de Baessa et al. n.d.). The discussion of methods says that 'several statistical techniques were used to analyse the data, such as two-tailed t-tests, analysis of covariance and chi square, depending on the type of data to be analysed in each case' (de Baessa et al. n.d., i). There is no elaboration on the exact statistical specifications or rationale for using a given technique. Complete statistical results are not reported.
14. The reported data contrast with the optimistic citation of the evaluations findings in secondary sources (Kraft 1998).
15. In fact, there are few such control variables available in early rounds of the SIMCE data. However, other authors have used a school-level index of socio-economic status developed by JUNAEB, the agency that administered the school meal programme (McEwan and Carnoy 2000).



## Notes on contributor

Patrick J. McEwan is an associate professor of economics at Wellesley College. His research interests include the economics of education, applied econometrics, and Latin American education systems. He has conducted evaluations of education reforms throughout the region.

## References

- Benveniste, L.A., and P.J. McEwan. 2000. Constraints to implementing educational innovations: the case of multigrade schools. *International Review of Education* 46, no.1–2: 31–48.
- Berry, C. 2001. Achievement effects of multigrade and monograde primary schools in the Turks and Caicos Islands. *International Journal of Educational Development* 21, no. 6: 537–552.
- . 2006. Learning opportunities for all: Pedagogy in multigrade and monograde classrooms in the Turks and Caicos Islands. In *Education for all and multigrade teaching: Challenges and opportunities*, ed. A.W. Little, 27–46 Dordrecht: Springer.
- Birdsall, N., R. Levine and A. Ibrahim. 2005. *Toward universal primary education: investments, incentives, and institutions*. London: UN Millennium Project, Earthscan.
- Centro de Investigación y Desarrollo de la Educación (CIDE). 2002. *Estudio dotación de recursos pedagógicos para estudiantes y docentes de las escuelas rurales adscritas al Programa de Educación Básica Rural* [Student and teacher access to pedagogical resources in rural schools participating in the Basic Rural Education Programme]. Santiago: CIDE.
- Chay, K.Y. P.J. McEwan and M. Urquiola. 2005. The central role of noise in evaluating interventions that use test scores to rank schools. *American Economic Review* 95, no. 4: 1237–1258.
- Chesterfield, R., and F.E. Rubio. 1997. *Estudio de impacto del Proyecto BEST en la efectividad de maestros en la educación básica de Guatemala* [Impact of the BEST Project on the effectiveness of basic education teachers in Guatemala]. Washington, DC: US Agency for International Development.
- Colbert, V., C. Chiappe, and J. Arboleda. 1993. The New School Program: more and better primary education for children in rural areas of Colombia. In *Effective schools in developing countries*, ed. H.M. Levin and M.E. Lockheed, 52–68 London: Falmer.
- Cook, T.D., H.D. Hunt and R.F. Murphy. 2000. Comer's School Development Program in Chicago: A theory-based evaluation. *American Educational Research Journal* 37, no. 2: 535–597.
- Cox, C. 2004. Innovation and reform to improve the quality of primary education: Chile. Unpublished manuscript, 2005 Education for All Global Monitoring Report.
- Craig, H.J., R.J. Kraft and J. du Plessis. 1998. *Teacher development: making an impact*. Washington, DC: US Agency for International Development and the World Bank.
- de Baessa, Y. n.d. *Informe final de investigación, Proyecto de mejoramiento de la calidad educativa, Guatemala* [Final research report, Education Quality Improvement Project, Guatemala]. Institute for International Research, Juárez and Associates, University of Pittsburgh.
- de Baessa, Y., R. Chesterfield and T. Ramos. 2002. Active learning and democratic behavior in Guatemalan rural primary schools. *Compare* 32, no. 2: 205–218.
- de Baessa, Y., R.J. Girón, T. Ramos and J. Valdés. n.d. *Improving educational quality project, research report, year two, Guatemala*. Arlington, VA: Institute for International Research.
- Forero-Pineda, C., D. Escobar-Rodríguez and D. Molina. 2006. Escuela Nueva's impact on the peaceful social interaction of children in Colombia. In *Education for all and multigrade teaching: challenges and opportunities*, ed. A.W. Little, 265–299 Dordrecht: Springer.
- García-Huidobro, J.E. 2000. Educational policies and equity in Chile. In *Unequal schools, unequal chances*, ed. F. Reimers, 161–178 Cambridge, MA: Harvard University Press.
- Glewwe, P., and M. Kremer. 2006. Schools, teachers, and education outcomes in developing countries. In *Handbook of the economics of education*, ed. E. Hanushek and F. Welch, 945–1017. Amsterdam: Elsevier.
- Herrera Lara, R.H. 1999. Cultural images, education and domination: The case of the Mapuche of southern Chile. Unpublished Ph.D. dissertation, University of Illinois at Urbana-Champaign.
- Juárez and Associates. 2003. *The effects of active learning programs in multigrade schools on girls' persistence in and completion of primary school in developing countries*. Juárez Associates, US Agency for International Development).
- Jimenez, E., and Y. Sawada. 1999. Do community-managed schools work? An evaluation of El Salvador's EDUCO program. *World Bank Economic Review* 13, no. 3: 415–441.

- Kraft, R.J. 1998. *Rural educational reform in the Nueva Escuela Unitaria of Guatemala*. Washington, DC: Academy for Educational Development.
- Laboratorio Latinoamericano de la Evaluación de la Calidad de la Educación (LLECE). 1998. *Primer estudio internacional comparativo sobre lenguaje, matemática y factores asociados en tercero y cuarto grado* [First comparative international study on language, mathematics and associated factors in the third and fourth grades]. Santiago: UNESCO, LLECE.
- Little, A.W. 2001. Multigrade teaching: Towards an international research and policy agenda. *International Journal of Educational Development* 21, no. 6: 481–497.
- Little, A.W., ed. 2006. *Education for all and multigrade teaching: challenges and opportunities*. Dordrecht: Springer.
- Loeb, S., and P.J. McEwan. 2006. An economic approach to education policy implementation. In *New directions in education policy implementation: confronting complexity*, ed. M. Honig, 169–186. Albany, NY: State University of New York Press.
- McEwan, P.J. 1998. The effectiveness of multigrade schools in Colombia. *International Journal of Educational Development* 18, no. 6: 435–452.
- McEwan, P.J. 2001. The politics of rural school reform: Escuela Nueva in Colombia. *Journal of Education Policy* 16, no. 6: 547–559.
- . 2004. The indigenous test score gap in Bolivia and Chile. *Economic Development and Cultural Change* 53, no. 1: 157–190.
- McEwan, P.J., and Carnoy, M. 2000. The effectiveness and efficiency of private schools in Chile's voucher system. *Educational Evaluation and Policy Analysis* 22, no. 3: 213–239.
- McEwan, P.J., and Trowbridge, M. 2007. The achievement of indigenous students in Guatemalan primary schools. *International Journal of Educational Development* 27, no. 1: 61–76.
- McGinn, N. 1996. Resistance to good ideas: Escuela Nueva in Colombia. Unpublished manuscript, 1996 Conference of the Nordic Association for the Study of Education in Developing Countries.
- McLaughlin, M.W. 1991. Learning from experience: Lessons from policy implementation. In *Education policy implementation*, ed. A.R. Odden, 185–195. Albany: State University of New York Press.
- Ministerio de Educación. n.d. *Programa de Educación Rural Básica* [Basic Rural Education Programme]. Santiago: Ministerio de Educación.
- Misión Social. 1997. La calidad de la educación y el logro de los planteles educativos [The quality of education and the achievement of schools]. *Planeación y Desarrollo* 28, no. 1: 25–62.
- Perfetti, M. 2004. *Estudio sobre la educación para la población rural en Colombia* [Education in rural areas of Colombia]. Santiago: FAO–UNESCO.
- Pitt, M.M., M.R. Rosenzweig, and D.M. Gibbons, 1993. The determinants and consequences of the placement of government programs in Indonesia. *World Bank Economic Review* 7, no. 3: 319–348.
- Programa de Promoción de la Reforma Educativa en América Latina y el Caribe (PREAL). 2003. *Desarrollo de la educación en sectores rurales*. Serie Mejores Prácticas. Santiago: PREAL.
- Psacharopoulos, G., C. Rojas and E. Velez. 1993. Achievement evaluation of Colombia's Escuela Nueva: Is multigrade the answer? *Comparative Education Review* 37, no. 3: 263–276.
- Rawlings, L.B., and G.M. Rubio. 2005. Evaluating the impact of conditional cash transfer programs. *World Bank Research Observer* 20, no. 1: 29–55.
- Rojas, C., and Z. Castillo. 1988. *Evaluación del Programa Escuela Nueva* [Evaluation of the New School programme]. Bogotá: Instituto Ser de Investigación.
- Schiefelbein, E. 1991. *In search of the school of the XXI century: is the Colombian Escuela Nueva the right pathfinder?* Santiago: UNESCO-UNICEF.
- Shadish, W.R., T.D. Cook and D.T. Campbell, 2002. *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
- UNESCO. 2004. Educación de la población rural: una baja prioridad [Rural education: A low priority]. *Boletín del Sector Educación de la UNESCO* 9, no 9: 4–7.
- Universidad Austral and Universidad de Playa Ancha. 1998. *Estudio de evaluación de la línea de educación rural del Programa MECE* [Evaluation of the rural education component of the MECE Programme]. Valdivia and Valparaíso: Universidad Austral and Universidad de Playa Ancha.
- Urquiola, M., and V. Calderón. 2006. Apples and oranges: educational enrollment and attainment across countries in Latin America and the Caribbean *International Journal of Educational Development* 26, no. 6: 572–590.

- Vithanapathirana, M. 2006. Adapting the primary mathematics curriculum to the multigrade classroom in rural Sri Lanka. In *Education for all and multigrade teaching: challenges and opportunities*, ed. A.W. Little, 127–154 Dordrecht: Springer.
- Vu, T.S., and P. Pridmore. 2006. Improving the quality of health education in multigrade schools in Vietnam. In *Education for all and multigrade teaching: challenges and opportunities*, ed. A.W. Little, 155–168 Dordrecht: Springer.
- Williamson, G. 2004. *Estudio sobre la educación para la población rural en Chile* [Education in rural areas of Chile]. Santiago: FAO–UNESCO.
- Winkler, D. 2000. Educating the poor in Latin America and the Caribbean: Examples of compensatory education. In *Unequal schools, unequal chances*, ed. F. Reimers, 113–132. Cambridge, MA: Harvard University Press.
- World Bank. 2005. *Central American education strategy: an agenda for action*. Washington, DC: World Bank.
- World Education Forum. 2000. *Education for all: meeting our collective commitments*. UNESCO. [www.unesco.org/education/efa/ed\\_for\\_all/dakfram\\_eng.shtml](http://www.unesco.org/education/efa/ed_for_all/dakfram_eng.shtml) (accessed 17 May 2006).