Programa de Promoción de la Reforma Educativa en América Latina y el Caribe
Partnership for Educational Revitalization in the Americas

No. 38

The Costs of Student Assessments in Latin America

By Laurence Wolff

September 2007

For additional copies, contact PREAL at the Inter-American Dialogue. The report can be downloaded from www.preal.org/publicacion.asp.

Inter-American Dialogue
1211 Connecticut Avenue, NW, Suite 510
Washington, DC 20036
202-822-9002
iad@thedialogue.org
www.thedialogue.org and www.preal.org

Citation: Wolff. Laurence, *The Costs of Student Assessments in Latin America* (Washington, DC: PREAL, 2007).

ISBN: 09772271-7-0
First edition
Published in the USA

Book design: Nita Congress
Cover design: Studio Grafik

Contents

List of Tables

List of Figures

Acknowledgments

The author would like to thank officials and researchers Lorena Meckes (Chile), Daniel Bogoya and Álvaro Díaz (Colombia), Renán Rápalo (Honduras), Liliana Miranda (Peru), and Pedro Ravela and Beatriz Picaroni (Uruguay) for their kindness in providing the information that went into this report and for commenting on drafts; Lucrecia Palacios and Julien Hautier for assistance in data analysis; Jeff Puryear, Hans Wagemaker, and Andreas Schleicher for comments on various drafts; and Kristin Saucier for putting the document into final form. The author is extremely grateful to Patricia Arregui for her continuous support in the initiation and preparation of this report. Any mistakes and inconsistencies in the data are the responsibility of the author.

PREAL's activities are made possible by the generous support of the American people through the United States Agency for International Development (USAID), by the Inter-American Development Bank (IDB), the GE Foundation, the International Association for the Evaluation of Educational Achievement (IEA), and the World Bank, among others. However, the contents of this document are the responsibility of the author and do not necessarily reflect the views of PREAL or any of its donors.

Acronyms and Abbreviations

ANEP	Administración Nacional de la Educación Pública (National Administration for Public Education – Uruguay)
CBEEE	curriculum-based external exit exam
GDP	gross domestic product
ICFES	Instituto Colombiano Para el Fomento de la Educación Superior (Institute for the Development of Higher Education – Colombia)
IDB	Inter-American Development Bank
IEA	International Association for the Evaluation of Educational Achievement
LLECE	Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (Latin American Laboratory for the Measurement of Quality in Education)
PIRLS	Progress in Reading and Literacy Study
PISA	Programme for International Student Assessment
PPP	purchasing power parity
SABER	Sistema Nacional de Evaluación de la Calidad de la Educación (National System for Evaluating Educational Quality – Colombia)
SIMCE	Sistema de Medición de la Calidad de la Educación (National System for the Assessment of Educational Quality – Chile)
TIMSS	Trends in International Mathematics and Science Study
UMC	Unidad de Medición de la Calidad Educativa (Educational Quality Measurement Unit – Peru)
UMCE	Unidad de Medición de la Calidad Educativa (Unit for Measuring Educational Quality – Honduras)
UNESCO	United Nations Educational, Scientific, and Cultural Organization

Executive Summary

Testing, at least on a sample basis, is rapidly becoming a fundamental element in modern education systems, both because it is a necessary part of the process of designing, implementing, and evaluating programs to improve the quality of education and because countries around the world are increasingly testing all students, usually in selected school years. Up to now there has been little published information on the subject of the costs of testing, and none for Latin America. The purpose of this report is to provide preliminary estimates on the subject, based on information provided by Chile, Colombia, Honduras, Peru, and Uruguay. Each of these countries has followed its own set of criteria and definitions, and field work was not undertaken to confirm their estimates. In particular, overhead costs of managing a testing unit are not included. The data presented should therefore be considered indicative rather than definitive. Data were available on censal testing in Chile and Colombia; on sample tests in four of the five countries; and on the costs of participating in international (sample-based) tests in Chile, Colombia, and Peru.

This review yielded the following findings:

- Costs vary greatly from one country to another even when roughly the same number of students are tested. Several reasons account for this variation, such as the following:
  - Some countries test multiple grades and subjects at one time, and with broader and/or more in-depth curricular coverage, thus increasing costs.
  - Some countries collect information on possible determinants of differential achievement, while others do not.
  - Countries with large populations and land area are more costly to sample than more compact countries.
  - Open-ended questions are more costly to score than multiple-choice questions.
  - Countries with limited human resources may need to utilize more expensive local and foreign consultants.
  - Some countries consider it desirable to have teachers and other sector officials administer the tests, while others prefer to hire external test administrators.

— Wage scales may vary from country to country even after accounting for purchasing power parity (PPP).

Cost comparisons—which are not the main thrust of this study—should take these differences into account.[1]

- The absolute costs of testing vary from US$111,000 (PPP) in Uruguay for a sample test to $6.5 million in Chile for a test of all eighth graders.[2] Understandably, censal testing is more expensive than sample testing. The costs of participating in international testing programs such as the Programme for International Student Assessment and the Trends in International Mathematics and Science Study ranged from $311,000 to $599,000 across three countries.

- Per student costs obviously vary greatly depending on the number of students tested. They range from $2.50 per student in testing over a million students in Colombia to around $100 per student in Peru and Colombia for testing a sample of 5,000 for an international test.

These many differences notwithstanding, the costs of testing, as currently practiced in the region (e.g., sample surveys or censal tests of selected grades), *are not an overly heavy burden on education budgets*. In none of the countries studied does testing involve more than 0.3 percent of the national education budget at the level (primary or secondary) tested.

On the basis of the cost data provided, as well as a review of the literature on the impact of testing on learning achievement, the following conclusions may be drawn:

- Testing is among the least expensive innovations in primary education reform, costing far less than increasing teachers' salaries, reducing class size, and reforming teacher training.

- Costs play an important but not defining role in decisions about testing. Each country has a different set of conditions, and decision makers and technicians need to make their own trade-offs regarding breadth and depth of testing based on their objectives and capacities. Given current capacities, it is not advisable to test all students in all grades, as is now mandated in the United States.

---

[1]The tables in the appendix are provided to facilitate such comparisons.

[2]All monetary values in this report are expressed in equivalent U.S. dollars, converted using PPPs.

- The only situations in which testing can be considered a poor use of public funds is when the technical quality of the tests is so low as to preclude drawing any valid conclusions about learning, or when the information gathered from the tests is neither disseminated nor used. Decision makers should not underestimate the complexity and technical challenges of measuring learning achievement and should make every effort to ensure high technical quality. Additionally, a fully funded dissemination plan should be in place before starting any testing program.

- There is increasing evidence around the world that "high-stakes testing," or curriculum-based external exit exams designed to certify that students have successfully completed a level of education (usually secondary education), can increase the level of learning achievement. Countries may wish to look at this option as a means of increasing learning achievement in secondary education.

- Participating in international tests is not expensive, and it can pay off many times over if the results are used to reform curricula and teacher training.

- More money should be spent on measuring what "works" in education, including, but not restricted to, the possible impact of testing on learning achievement.

I. Why Estimate the Costs of Testing?

The main purpose of this report is to provide information on the costs of testing in Latin America. Until now, there has been no published information on this topic. There have been anecdotal complaints about the high costs of testing, especially as compared to the limited discretionary funds available to ministries as well as the extensive time demands on small numbers of qualified personnel. The debate on the costs of testing appears to be "under the radar," and not yet within public awareness. This may change as testing becomes more widespread, and as the external agencies currently financing testing in a number of countries insist that these costs be integrated into regular budgets.

Testing Is Fundamental

This report assumes that testing, at least on a sample basis, is a fundamental element in any modern education system, since it is a necessary part of the process of designing, implementing, and evaluating programs to improve the quality of education. It is based on data graciously made available by officials from five countries in Latin America—Chile, Colombia, Honduras, Peru, and Uruguay—on three types of tests:

- universal or censal testing (that is, tests applied to all students in selected grades) in Chile and Colombia;
- national sample testing (that is, tests applied to random samples of students so as to generalize results for the nontested universe) in Colombia, Honduras, Peru, and Uruguay; and
- participation in international programs of sample testing in Colombia, Peru, and Uruguay.

Those who are in charge of their national testing programs or their designated colleagues provided the information for this study. Costs are compared with the costs of educating students at the level in which the test was given and as a percentage of overall per student costs in each country. Costs are presented in local currencies and in U.S. dollars, as estimated in "purchasing power parity" (PPP) in the year in which a test was given.[1] Each country reporting this information followed its own set of criteria and definitions; no field work was undertaken to confirm their estimates. The data presented should therefore be considered indicative rather than definitive. It is expected that future studies will more finely gauge the overall costs of assessments in these and, hopefully, many other countries.

Three Basic Types of Tests

In Latin America, 16 countries now test their students either on a census or sample basis. Chile regularly tests all students in selected grades. Other entities in the region that test all students in selected grades and years include Mexico, Colombia, and several states in Brazil. El Salvador has recently begun testing all students in selected grades, while Guatemala is planning to do so in the near future. These countries use censal testing for a variety of purposes, but especially for diagnosis and feedback to stakeholders. Chile uses the data to identify schools and localities that need

[1]This ratio, used by the World Bank, the Organisation for Economic Co-operation and Development, and other international agencies, is designed to bypass issues of over- or undervaluation of currencies based on official exchange rates by taking into account the purchasing power of local currencies and thus making the numbers more nearly comparable. Since currencies in most developing countries are undervalued, PPP estimates usually show higher U.S. dollar estimates compared with estimates based on official exchange rates. For example, per capita income in Chile and Peru is over twice as high using PPP exchange rates compared to official exchange rates; per capita income in Colombia is three times higher; and in Honduras, it is 2.5 times higher. PPP and official exchange rates for 2002 can be found at the World Bank Web site at http://siteresources.worldbank.org/ICPINT/Resources/Table5_7.pdf. All monetary values in this report are expressed in equivalent U.S. dollars converted using PPPs.

help, reward well-performing teachers, study causes of failure or success, and inform parents about their children's progress. In Chile and Mexico, relative scores of students on tests have an impact on teachers' salaries or promotion possibilities.

Censal testing can be used to determine whether students will be able to receive a diploma and/or move on to the next level of education. These "high-stakes" tests are normally given to all students at the end of a cycle, usually secondary education. They are common in Europe, the Far East, and former Anglophone and Francophone countries, including the English-speaking Caribbean, as well as in a few states in the United States. In Latin America, Costa Rica, the Dominican Republic, and El Salvador now require a secondary school test that accounts for between 20 and 25 percent of a student's grades at the end of secondary school. English-speaking Caribbean countries test all students in a high-stakes context (usually, but not solely, upon secondary school completion).

Tests may also be regularly given to samples of students, as is the practice of most of the countries in the region that have testing programs. Sample testing (as well as censal testing) may be used to determine the extent to which the national curriculum and agreed-upon standards are being successfully implemented in the classroom, as well as to study what factors are contributing to differential results of various student populations. They may help in designing training programs and materials for teachers in areas where large percentages of students do poorly, and the data can be used to assess the impact of various programs and policies. However, with sampling, it is not possible to identify, reward, or sanction individual schools or teachers for their relative performance. In Peru, relatively large samples have sometimes been tested to get an idea of progress in subnational territorial units; in Colombia, results are available that allow finer analysis of what is happening within a metropolitan area. Uruguay used sampling to evaluate the impact of a teacher training reform, introduction of full-day schooling, and programs directed at children from poorer families. In Honduras, sampling has been used to get an overall idea of learning in the country, to compare learning by region and town, and to identify factors associated with learning. The federal government of Brazil regularly samples students to compare progress in states and municipalities.

A variation on sample testing occurs when countries participate in international testing programs, as is increasingly happening in Latin America (see box). Since the purpose of these tests is to place a country in the context of other countries, a relatively small

Recent Latin American Participation in International Tests

- Latin American Laboratory for the Measurement of Quality in Education (Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación—LLECE): 13 countries in 1997; 16 countries plus one state in Mexico in the second round in 2006
- The Organisation for Economic Co-operation and Development's Programme for International Student Assessment (PISA): Argentina, Brazil, Chile, Mexico, Peru, and Uruguay in 2002 and/or 2003; Argentina, Brazil, Chile, Colombia, Mexico, and Uruguay in 2006
- Trends in International Mathematics and Science Study (TIMSS) of the International Association for the Evaluation of Educational Achievement (IEA): Argentina, Chile, Colombia, and Mexico in 2003
- Progress in Reading and Literacy Study (PIRLS), another IEA-supported initiative: Argentina, Belize, and Colombia in 2001

random sample—around 5,000 to 7,000 students—is usually required. This is the minimum to ensure a reasonable confidence interval for the results. In some cases, samples of this size are taken in each of several provinces so that each one is in effect considered as a country. In PISA 2003 and 2006, Mexico implemented a sample of more than 30,000 students in order to get reliable data for each state. Helping establish new national goals for learning achievement, or reforming curriculum and/or teacher training so as to achieve such goals, are possible objectives for participating in international tests. A full summary of the testing situation in Latin America is provided by Ferrer (2006).

The choice of whether to apply censal or sample testing needs to be made on the basis of carefully thought-out and planned uses of test results—something that has been frequently overlooked in Latin America (see Ravela 2001, "Introduction"). Table 1 provides a summary of the potential objectives of testing, organized by type of test.

Study Overview

As we shall see, the costs per student differ significantly from one country to another, but this should not be interpreted to mean that one country is spending "too much" per student on testing, but rather that the conditions for testing differ greatly from one country to another. It's far easier and less costly, for example, to undertake a random sample in Uruguay than in Peru: the latter's huge physical size and varied and difficult terrain significantly increase the costs (notably of transportation) of undertaking a random sample. Similarly, a test with only multiple-choice questions, as is used in Colombia, is far less expensive to correct

Table 1: Potential objectives of tests conducted in Latin America

Type of test	Characteristic	Potential objective
Participation in international testing	Sample, usually of about 5,000 to 7,000 students	1. Catalyze national debate on improving learning and education reform 2. Assess adequacy of intended and implemented curriculum in the classroom in terms of its relative emphasis on higher order learning, reading comprehension, etc. 3. Assess adequacy of in-service and preservice teacher training and other educational inputs, including school organization and teaching practices
National sample	Sample, ranging from 5,000 to 100,000 students	All of the above, as well as 4. Track progress over time of country, regions, and specific populations toward specified learning goals 5. Compare performance of regions and various population groups (urban/rural, male/female, ethnic identity, socioeconomic status) in an effort to find explanations for differences in achievement 6. Target districts or population groups for improvement efforts, rewards, and/or sanctions
Censal testing of all students in a grade	Universal, usually of selected grades and undertaken in selected years, of up to 1 million or more students, depending on the size of the country	All of the above, as well as 7. Provide feedback to students and parents on individual student progress in learning 8. Provide feedback to teachers, schools, and parents on school and teacher performance and offer rewards and sanctions to stimulate improvement efforts 9. For high-stakes exams, certify that students have completed a level of education and/or are eligible for higher levels

than one with many open-ended questions. The purpose of this report is not to compare high or low costs among countries but rather to give an overview of the range of costs of testing in the region and to identify situations on the ground, options, and trade-offs that affect costs. A more rigorous determination of costs to allow for more precision and comparability among countries is a challenge that lies ahead and will entail the tracking of costs by countries, beginning with the first stages of testing.

For this report, officials from five countries took the time to retrieve and organize data on costs using a standard spreadsheet provided to them. In some cases, countries did not provide all the data requested, or they collapsed them into several categories. Overhead costs were available in some cases and not in others. Furthermore, some hidden costs, such as the time teachers and supervisors spend administering tests, were not estimated.

Countries provided data in four general categories:

- **Test preparation:** analysis of curriculum and determination of the education objectives to be tested, preparation of items and the first draft of the test, pilot testing and feedback, and preparation of the final version of the test.
- **Test application:** distribution of material; preparation of the team to supervise the test, including contracting of supervisors, testers, and proctors; and data collection.
- **Processing and analysis:** training personnel; processing of multiple-choice questions, usually by automatic data processing; correction of open-ended questions; analysis of results; and preparation of reports.
- **Dissemination:** printing and dissemination of documents and reports.

The estimates provided usually do not include training of teachers, curriculum development, or textbook revision based on test results, but may include the costs of short-term seminars.

The next section of this report summarizes the available information on costs from the five countries providing data—namely, Chile, Colombia, Honduras, Peru, and Uruguay. As noted earlier, costs depend fundamentally on the characteristics of the populations being sampled and the breadth and depth of the assessment being attempted. Thus, readers should review the appendix, which summarizes the characteristics of the five national cases used in this study.

Section III draws tentative conclusions regarding testing in the region, with a focus on the three types of tests—national censal tests, sample-based tests, and international tests. The final section discusses the potential impact of testing on learning compared to its costs.

II. Costs of Testing in Latin America

Censal Testing

Data were made available from two countries that test all students in particular grades. Chile provided data on eighth grade tests given in 2004 for four academic subjects. Colombia provided data on tests given in 2003 and 2004 for grades 5 (language and math) and 9 (science and citizenship).

Chile

Chile has been administering SIMCE—the National System for the Assessment of Educational Quality (Sistema de Medición de la Calidad de la Educación)—for over 15 years, and so has developed expertise and a cadre of experts both within and outside govern-

ment. The SIMCE tests are given approximately every four years to grades 4, 8, and 10.[2] The 2004 test was given to 300,000 eighth grade students in 6,500 schools and 11,000 classrooms in the subject areas of Spanish, mathematics, the natural sciences, and the social sciences. The test included a combination of multiple-choice and open-ended questions.

The data provided by Chile are much more detailed than those provided by other countries in the sample. Notably, Chile was able to break down the activities in test preparation, application, and analysis. Chile was able to separate out the costs of correction of multiple-choice questions (57 percent of the total cost of processing and analysis) from correction of open-ended questions (43 percent). Also, Chile (like a few other countries) provided some information on overhead costs—specifically, the cost of salaries for those in the assessment unit who oversee all testing activities. Table 2 shows costs by category in Chilean pesos as well as in U.S. dollars using PPP exchange rates.

SIMCE prepared a detailed description of how costs were calculated for each line item, as follows:

- **Elaboration of items:** contracted with a university, 200 validated items, including guidelines for marking open-ended questions.
- **Initial piloting:** a sample seeking 300 valid student responses per item.
- **Pilot testing:** a sample seeking 3,000 valid student responses per item, including design, printing, and distribution.
- **Design and editing:** design of all materials, including test booklets, list of homerooms, questionnaires, forms, etc.
- **Test printing:** 300,000 test books for four learning areas, for a total of 1.2 million test books.
- **Printing of other materials:** homeroom lists, survey of parents and teachers, for a total of 400,000 documents.
- **Distribution**: to approximately 6,000 schools, of which about 25 percent were rural.
- **Field testing:** contracting of 10,000 examiners.
- **Control and supervision:** undertaken by departmental staff; costs include only travel and per diem.
- **Coding and digital input:** coding using mark sense scanners of all items and questionnaires.
- **Marking open-ended questions:** marking of all of two sections and 10 percent of two other sections, with double marking for 10 percent of answers.
- **Report to schools:** personalized results to each school; does not include other forms of communication.
- **Overhead—personnel:** salaries of 60 staff members.
- **Overhead—infrastructure:** physical space for SIMCE personnel.
- **Overhead—equipment:** computers.
- **Overhead—other:** general costs, including phone, electricity, etc.

The total cost for this particular test administration was about $6.5 million, or $14.90 per student tested in grade eight. If the full overhead costs of running the SIMCE program—mainly salary expenses for SIMCE staff involved with management and oversight—were added, then the cost per student would increase by 45 percent to $21.65. However, because SIMCE staff are regularly involved in managing and planning for several tests at one time, this figure may be lower.

[2]Detailed explanations and results for SIMCE are available at www.simce.cl.

Table 2: Chile—SIMCE tests

Test parameters		
Year	2004	
Subject	Language, mathematics, natural sciences, social sciences	
Type of school	Public, private	
Grade	8	
Number of students tested	300,000	
Test costs	Pesos	US$ (PPP)
Test preparation	**175,200,000**	**607,000**
Elaboration of items	76,000,000	263,000
Initial piloting	20,000,000	69,000
Pilot testing	30,000,000	104,000
Guidelines for marking open-ended questions	49,200,000	170,000
Test application	**788,700,000**	**2,732,000**
Test design and editing	20,000,000	69,000
Test printing	220,000,000	762,000
Printing of other materials	160,000,000	554,000
Distribution	70,000,000	242,000
Field work	275,000,000	953,000
Control and supervision	43,700,000	151,000
Processing and analysis	**258,900,000**	**897,000**
Coding and digital input	146,400,000	507,000
Marking open-ended questions	112,500,000	390,000
Dissemination	**68,250,000**	**236,000**
Report to each school[a]	68,250,000	236,000
Subtotal	**1,291,050,000**	**4,472,000**
Overhead costs	**584,700,000**	**2,025,000**
Personnel	540,000,000	1,870,000
Infrastructure	24,000,000	83,000
Equipment	7,200,000	25,000
Other	13,500,000	47,000
TOTAL	**1,875,750,000**	**6,497,000**
Cost per student[b]	4,304	14.90
Cost of educating a student[c]	519,371	1,799
Cost of testing as % of total budget for one grade	0.83	
Cost of testing as % of total secondary education budget	0.17	

Source: SIMCE.

Notes: 2002 PPP exchange rate: 288.7 pesos = US$1.

[a]Costs for other documents, publications, and meetings were not available.

[b]Not including overhead.

[c]Cost per student at the secondary level in 2000. See Wolff and Gurría (2005).

Not counting overhead, the proportional breakdown of costs by activity is as follows: test preparation, 10.1 percent; test application, 63.5 percent; processing and analysis, 20.9 percent; dissemination, 5.5 percent. Within the category of test preparation, preparing and writing items (e.g., defining the domains to be tested and then writing items) accounted for 60 percent of costs. For the second category, test application, printing accounted for 26 percent of costs and field testing for 36 percent. Chile's inclusion of a background questionnaire for students, teachers, and some parents is the likely explanation for the relatively high costs of "other materials."

According to the United Nations Educational, Scientific, and Cultural Organization (UNESCO), the estimated cost of educating one student in the eighth grade in 2000 in Chile was $1,799.[3] Therefore, Chile's per student testing cost of $14.90 (not including overhead) would be equivalent to 0.8 percent of the cost of educating one eighth grader. Since secondary school consists of four grades and only one grade is tested in any given year, it could be said that the cost of secondary school testing is equivalent to one-quarter that amount, or 0.17 percent of total public expenditures on secondary education.

Colombia

Since 1980, Colombia's Institute for the Development of Higher Education (Instituto Colombiano Para el Fomento de la Educación Superior—ICFES), a public autonomous agency, has been testing all students who complete secondary education as a means of determining entrance to higher education institutions in the country; these assessments are similar to the entrance examinations of the Educational Testing Service of the United States.[4] Until recently, the national government of Colombia administered only sample surveys of learning in primary and lower secondary education.[5] In 2002 and 2003, it began testing all students in grades 5 and 9 every three years through its National System for Evaluating Educational Quality (Sistema Nacional de Evaluación de la Calidad de la Educación—SABER) program, which is also entrusted to ICFES. ICFES, which has been active in large-scale testing since the 1960s, has professional competencies and economies of scale often lacking in other countries.

Table 3 shows the estimated costs of universal student testing in Colombia, as reported by ICFES staff, for the testing of all students in fifth and ninth grades in 2002 and 2003. In 2002, students were tested in language and mathematics; and in 2003, in science and citizenship. For purposes of comparison, the costs of administering the higher education entrance examination (which is not an evaluation) are also included.

The data provided by Colombia are incomplete. For 2002, the costs of test processing and analysis and of dissemination were not estimated. For 2003, the costs of test processing and analysis were not estimated. Overhead costs were not estimated for either year.

In 2002, tests in mathematics and language were given to over 1 million students in grades 5 and 9 at a total cost of $2.5 million, or a per student cost (for students in these grades) of $2.47. This is substantially lower than Chile's $14.90 per student cost. As noted above, these figures do not include correction

[3]Most recent year available from international sources as of this writing.

[4]Information about Colombian tests can be found at www.colombiaaprende.edu.co/html/investigadores/1609/find-results.html and www.icfes.gov.co.

[5]The city of Bogotá has been testing all students in selected grades (third, fifth, seventh, and ninth) since 1998.

Table 3: Colombia—National evaluation of basic education and higher education entrance exam

Test parameters	2002		2003		2004	
Subject	Language, mathematics		Science, citizenship		Higher education entrance exam	
Type of school	Public, private		Public, private		Public, private	
Grade	5, 9		5, 9		11	
Number of students tested	1,030,626		1,034,049		540,716	
Test costs	**Pesos**	**US$ (PPP)**	**Pesos**	**US$ (PPP)**	**Pesos**	**US$ (PPP)**
Test preparation	113,081,000	157,000	29,039,000	38,000	795,127,000	1,002,000
Test application	1,710,842,000	2,380,000	4,009,667,000	5,298,000	5,878,419,000	7,408,000
Correction and analysis	8,133,000	11,000	[a]	[a]	[a]	[a]
Dissemination	NA	NA	121,850,000	161,000	800,126,000	1,008,000
TOTAL	**1,832,057,000**	**2,548,000**	**4,160,555,000**	**5,497,000**	**7,473,672,000**	**9,418,000**
Cost per student	1,778	2.47	4,024	5.32	13,822	17.42
Cost of educating a student[b]	1,089,948	1,365	1,147,364	1,365	1,203,704	1,516
Cost of testing as % of total cost of educating students in one grade[c]	0.2		0.4			
Cost of testing primary students as % of total primary education budget[d]	0.02		0.04			
Cost of testing secondary students as % of total secondary education budget[d]	0.02		0.05		0.17	

SOURCE: ICFES.

NOTES: NA = not available. Overhead costs are not included. 2002 PPP exchange rate: 719 pesos= US$1; 2003: 757 pesos = 1 US$1; 2004: 794 pesos = US$1.

[a]Included in test application.

[b]Weighted average of the cost per student at the primary and secondary levels in 2000, based on the number of students enrolled at each level. See Wolff and Gurría (2005).

[c]Total costs of testing in one grade as a percentage of total expenditures for that grade.

[d]Assumes that the number of students tested at each level correlates to the distribution of students in the respective level; data are for 2000.

and dissemination, which would likely increase the cost of testing by 25 percent. The impact of the costs of ICFES overhead cannot be estimated. Since Colombia spends an average of $1,365 to educate each of its students, its reported costs in 2002 are equivalent to 0.2 percent of the cost per student in each grade tested; and 0.02 percent of total expenditures in both primary and secondary education.

In 2003, Colombia tested all children in the areas of science and citizenship. The costs of implementing these tests was $5.32 per student tested, or more than double that of the tests in language and mathematics. Surprisingly, less than 1 percent of costs went to preparation and 3 percent to dissemination.

As noted above, Colombia has had a long-standing program of testing students in the last year of secondary education; this originally applied only to those who wish to continue in higher education. Up to nine subjects are tested, and about 500,000 students take these tests annually. The estimated cost is around $17.42 per student, which is significantly higher than the tests in primary and lower secondary education. The secondary school exit exams are 100 percent multiple choice.

There are a number of possible reasons that the reported costs of testing in Colombia are less than half those of Chile. Because ICFES staff code and analyze responses themselves—unlike what happens in Chile, where reliance on consultants is greater—it is difficult to determine this cost. Also, ICFES exclusively uses multiple-choice items rather than a combination of open-ended and multiple-choice items. ICFES tests do not include detailed background questionnaires. It is also possible that PPP exchange rates do not fully capture relative differences in salaries between the countries. The large number of students tested may give Colombia economies of scale not available to Chile. In addition, in 2002, ICFES implemented a system of rationalization of costs which included training, computerization of all processes, decentralization, and printing of most materials in it own printing facilities. Further review and information would be needed to confirm the relative weight of each of these factors.

Sample Testing

Data on sample testing were available for Colombia, Honduras, Peru, and Uruguay; the results are described below. In reviewing this material, note that Peru is similar in size and population to Colombia; while Honduras, although similar in size to Uruguay, is a much poorer country.

Colombia

In 2004, Colombia undertook a national sample test of progress in mathematics and language for grades 5 and 9. Table 4 provides a summary of costs provided by ICFES. Information on analysis costs was not available from Colombia.

A total of 96,000 students were tested in grades 5 and 9, or less than 10 percent of the total number of students in these grades. Colombia sampled a relatively large number of students so that it could compare learning across cities and regions. The total costs for this sample were equivalent to 40 percent of the costs of censal testing undertaken only two years earlier—$1.4 million, compared to $2.5 million two years earlier to test 1 million students. To test each student, it cost $14.20 in the sample assessment, compared with $2.47 per student in the censal assessment. On the other hand, since only a sample of students were tested, the cost of the testing program as a percentage of the overall budget was 0.01 percent, or less than half that of the censal testing undertaken two years earlier. The proportional breakdown of costs on the sample test administration was as follows: test prep-

Table 4: Colombia—Sample testing of basic education

Test parameters		
Year	2004	
Subject	Language, mathematics	
Type of school	Public, private	
Grade	5, 9	
Number of students tested	96,242	
Test costs	**Pesos**	**US$ (PPP)**
Test preparation	100,000,000	126,000
Test application	838,318,000	1,056,000
Processing and analysis	NA	NA
Dissemination	159,414,000	181,000
TOTAL	**1,097,732,000**	**1,363,000**
Cost per student	11,405	14.20
Cost of educating a student[a]	1,203,019	1,365
Cost of testing primary students as % of total primary education budget[b,c]	0.01	
Cost of testing secondary students as % of total secondary education budget[c]	0.01	

SOURCE: ICFES.

NOTES: NA = not available. Overhead costs are not included. 2004 PPP exchange rate: 794 pesos = US$1.

[a]Weighted average of the cost per student at the primary and secondary levels in 2000, based on the number of students enrolled at each level. See Wolff and Gurría (2005).

[b]Calculated based on total costs of testing in one grade as a percentage of total expenditures on primary education, and not on the per student cost of testing, since only a sample of students were tested.

[c]Assumes that the number of students tested at each level correlates to the distribution of students in the respective level; data are for 2000.

aration, 9 percent; test application, 78 percent; test analysis, not provided; and dissemination, 13 percent.

Uruguay

Data are available from Uruguay's Administración Nacional de la Educación Pública (ANEP) on sample testing in 2001, 2002, and 2003 of preschool and 1st, 2nd, 6th, and 12th grade students. In 2002, Uruguay tested all those schools operating a new program of full-day schooling, as well as a random sample of other schools, in the areas of language and mathematics; a total of 9,171 students were tested. In 2003, it tested 12,993 students in "diversified" or upper secondary school—a sample of general schools and all students in technical schools—in the areas of language, mathematics, the social sciences, and the natural sciences. It also administered background questionnaires to students, teachers, school principals, and parents.[6] The reported costs are shown in table 5.

Uruguay's total costs for testing were $111,000 in 2002 and $266,000 in 2003. The per student cost in 2002 for sixth graders taking tests in two subjects was $12. The cost per student in 2003 was higher, $20; this is perhaps because four, rather than two, subjects were tested, and because testing 12th graders requires complex measures of higher order skills compared to sixth graders. The per student cost for the 2001 test was

[6]Information about Uruguay's testing program can be found at www.anep.edu.uy/.

Table 5: Uruguay—Sample testing of basic and secondary education

Test parameters	2001		2002		2003	
Subject	Cognitive and affective development language, mathematics		Language, mathematics		Language, mathematics, natural science, social science	
Type of school	Public, private		Public, private		Public, private	
Grade	Preschool, 1, 2		6		12	
Number of students tested	2,387		9,171		12,993	
Test costs	Pesos	US$ (PPP)	Pesos	US$ (PPP)	Pesos	US$ (PPP)
Test preparation	299,000	35,000	229,000	23,000	608,000	54,000
Test application	464,000	55,000	524,000	52,000	1,618,000	144,000
Correction and test analysis	36,000	4,000	134,000	13,000	742,000	66,000
Dissemination	64,000	8,000	226,000	23,000	15,000	2,000
TOTAL	**863,000**	**102,000**	**1,113,000**	**111,000**	**2,983,000**	**266,000**
Cost per student	362	43	121	12	230	20
Cost of educating a student	8,592	1,016	10,104	1,011	13,657	1,219
Cost of testing as % of total expenditures on primary education[a]	0.03		0.03			
Cost of testing as % of total expenditures on secondary education					0.07	

SOURCE: ANEP.

NOTES: PPP exchange rates varied for each year; in 2003, the rate was 11.21 pesos = US$1. Costs for the small number of full-time staff working on all studies were not included, nor were the costs of various dissemination activities undertaken by the staff and other institutions.

[a]Calculated based on total costs of testing in one grade as a percentage of total expenditures on primary education, and not on the per student cost of testing, since only a sample of students were tested.

even higher because of the small sample and the fact that three grades were tested. Since Uruguay has a far smaller education system and budget than Colombia and was testing a higher percentage of its students, Uruguay's cost as a percentage of its national budget was much higher than that of Colombia—0.03 percent in 2001 and 2002 and 0.07 percent in 2003, compared to 0.01 percent. The 2002 breakdown of costs by type of activity was as follows: preparation, 10.1 percent; application, 46.9 percent; processing and analysis, 11.7 percent; and dissemination, 20.7 percent. Uruguay paid a great deal of attention to dissemination of its 2002 test.

Peru

In 2001, Peru administered tests in mathematics and communication (reading and writing assessment) to 34,000 students in the fourth and sixth grades of primary school and the fourth grade of secondary school. In 2004, it undertook a very ambitious regimen, testing a total of 70,000 students in primary school grades 2 and 6 and secondary school grades 3 and 5, or about 17,500 students in each grade. The subjects tested in 2004 were mathematics and communication in all grades, and citizenship in primary school grade 6 and secondary school grade 5. The sixth grade language test was administered in

two indigenous languages as well as Spanish. Table 6 summarizes the test program costs.

The total costs of testing in 2001 were $1.7 million; Peru's 2004 costs were more than twice that amount: $4.9 million. In 2004, a third subject—citizenship—was included, more than twice the number of students were tested, and field control to ensure adequate and reliable testing conditions was reported to be much more rigorous. The average cost of testing per student was $52 in 2001 and $70 in 2004.

Peru's tests, especially those given in 2004, were comprehensive and complex. Four grades and three subjects were tested, compared to one or two grades in the case of the other countries providing data. Tests were also given in indigenous languages, with a special sample drawn from bilingual schools. Background questionnaires were given to students, parents, and teachers. Teachers were also given questionnaires on "opportunity to learn," which asked them what areas in language and mathematics they were actually teaching, and were tested on their own knowledge in these subjects. Test application and correction were contracted out. On the

Table 6: Peru—Sample testing of basic and secondary education

Test parameters	2001		2004	
Subject	National sample: mathematics, reading comprehension, writing Bilingual schools sample: reading comprehension and writing in first (Aymara and Quechua) and second (Spanish) languages		National sample: citizenship, mathematics, reading comprehension, writing Bilingual schools sample: reading comprehension and writing in first (Aymara and Quechua) and second (Spanish) languages	
Type of school	Public, private		Public, private	
Grade	4, 6, 10		2, 6, 9, 11	
Number of students tested	34,000		70,000	
Test costs	**Soles**	**US$ (PPP)**	**Soles**	**US$ (PPP)**
Test preparation	250,000	166,000	1,142,000	760,000
Test application	1,474,000	980,000	5,083,000	3,384,000
Processing and analysis	587,000	390,000	899,000	598,000
Dissemination	331,000	220,000	214,000	142,000
TOTAL	**2,642,000**	**1,757,000**	**7,338,000**	**4,885,000**
Cost per student	78	52	105	70
Cost of educating a student[a]	751	500	751	500
Cost of testing as % of total expenditures on primary education[b]	0.06		0.15	
Cost of testing as % of total expenditures on secondary education	0.05		0.13	

Source: UMC.

Notes: Overhead costs are not included. 2001/2004 PPP exchange rate: 1.5 soles = US$1.

[a]Weighted average of the cost per student at the primary and secondary levels in 2000, based on the number of students enrolled at each level. See Wolff and Gurría (2005).

[b]Calculated based on total costs of testing in one grade as a percentage of total expenditures on primary education, and not on the per student cost of testing, since only a sample of students were tested.

tests, open-ended questions accounted for between 30 and 100 percent of the items, and were scored twice to ensure reliability. The test given to the teachers was 100 percent open ended. Test administration lasted five days. Four- or five-day training courses for test administrators were provided. Peru's Educational Quality Measurement Unit (Unidad de Medición de la Calidad Educativa—UMC) managed test administration as its main activity. The World Bank and the Inter-American Development Bank (IDB) provided partial financing to the initiative. The total cost of testing in 2004—$4.9 million—was equivalent to 0.15 percent of the country's total expenditures on primary education and 0.13 percent of its expenditures on secondary education.

Honduras

Honduras tested over 40,000 third and sixth graders in 2002 and 2004 in language, mathematics, and science achievement.[7] Table 7 provides a summary of the costs, as provided by consultants in the Unit for Measuring Educational Quality (Unidad de Medición de la Calidad Educativa—UMCE) located at the Pedagogical University of Honduras.

Honduras tested between 24,000 and 25,000 students in grade 3 and between 18,000 and 20,000 students in grade 6 of the estimated 140,000 students in each grade. The total costs were $1.9 million in 2002 and $2.3 million in 2003. The costs per student tested were about $46 in 2002 and just over $50 in 2004. The total testing costs were equivalent to between 0.28 percent and 0.33 percent of the national budget expended on primary education. This is a significant amount, considering that over 90 percent of this budget goes to

[7]Information about Honduran testing can be found at www.upnfm.edu.hn/proyectos/umce.htm.

Table 7: Honduras—Sample evaluation

Test parameters	2002		2004	
Subject	Language, mathematics, natural science		Language, mathematics, natural science	
Type of school	Public, private		Public, private	
Grade	3, 6		3, 6	
Number of students tested	42,572		45,657	
Test costs	**Lempira**	**US$ (PPP)**	**Lempira**	**US$ (PPP)**
Test preparation	2,377,000	391,000	3,081,000	458,000
Test application	5,942,000	977,000	7,703,000	1,146,000
Processing and analysis	1,783,000	293,000	2,311,000	344,000
Dissemination	1,783,000	293,000	2,311,000	344,000
TOTAL	**11,885,000**	**1,953,000**	**15,406,000**	**2,292,000**
Cost per student	279	45.88	337	50.20
Cost of educating a student[a]	2,080	342	2,299	342
Cost of testing as % of total expenditures on primary education[b]	0.28		0.33	

SOURCE: UMC.

NOTES: Overhead costs are not included. 2002 PPP exchange rate: 6.1 lempiras = US$1; 2004: 6.6 lempiras = US$1.

[a]Weighted average of the cost per student at the primary and secondary levels in 2000, based on the number of students enrolled at each level. See Wolff and Gurría (2005).

[b]Calculated based on total costs of testing in one grade as a percentage of total expenditures on primary education, and not on the per student cost of testing, since only a sample of students were tested.

teachers' salaries, and given the relatively small numbers of students tested. The proportional breakdown of expenditures by activity in 2004 was as follows: preparation, 20 percent; application, 50 percent; processing and analysis, 15 percent; and dissemination, 15 percent. Honduras has a very small cadre of qualified testing experts; this shortage probably led to a heavier reliance on external consultants, thus driving up costs. Other elements of the Honduras program that had an effect on costs were the testing of multiple grades (two) and subjects (three), a more careful sampling and data analysis compared to that undertaken in the past, the procedural regulations dictated by the external funders, and the location of the assessment unit in the Pedagogical University rather than within the Ministry of Education, which was done, in part, to procure higher technical quality.

Participation in International Tests

Latin American countries are increasingly participating in international testing programs. Six countries (Argentina, Brazil, Chile, Mexico, Peru, and Uruguay) have participated in the PISA international testing effort, which measures reading and mathematical and science reasoning of 15-year-olds. The test includes a combination of open-ended and multiple-choice questions as well as a background questionnaire. PISA requires a minimum of at least 5,000 students randomly selected from an entire country to ensure statistically valid results. In 2003, it required each country to pay over $100,000 to be part of the program. Cost data on PISA were available from Uruguay for 2003 and for Peru for 2000. Cost data are also available from Colombia's participation in PIRLS, which measures the reading abilities of fourth graders, and which normally requires payment of $20,000 per year over three years from participating countries.

Uruguay

The costs of Uruguay's participation in the 2003 PISA are summarized in table 8.

Table 8: Uruguay—Participation in PISA

Test parameters		
Year	2003	
Subject	Mathematics, problem solving, reading, science	
Type of school	Public, private	
Grade	15-year-olds (mainly 10th graders)	
Number of students tested	5,797	
Test costs	**Pesos**	**US$ (PPP)**
Cost of participation	1,220,000	110,000
Test preparation	349,000	35,000
Test application	839,000	75,000
Processing and analysis	645,000	58,000
Dissemination	398,000	34,000
TOTAL	**3,451,000**	**311,000**
Cost per student	595	53.64
Cost of educating a student[a]	13,530	1,219
Cost of testing as % of total secondary education budget	0.08	

SOURCE: ANEP.

NOTES: 2003 PPP exchange rate: 11.21 pesos = US$1. Costs for the small number of full-time staff working on all studies were not included, nor were the costs of various dissemination activities undertaken by the staff and other institutions.

[a]Cost per student at the secondary level in 2000. See Wolff and Gurría (2005).

Uruguay's total costs for participation amounted to $311,000. The largest share of total expenditures was payment of $110,000 to PISA headquarters for central management, which included services in test preparation, guidance in field testing, international trips, and analysis. Uruguay tested a total of 5,797 students. Since the time of testing, Uruguay has published a number of analyses based on PISA; these may not have been included in the original dissemination costs. The pro-

portional breakdown of costs by testing activity was: participation in PISA, 35 percent; test preparation, 11 percent; test application, 24 percent; processing and analysis, 19 percent; and dissemination, 11 percent.

Peru

Table 9 summarizes the costs of Peru's participation in PISA 2000.

Table 9: Peru—Participation in PISA

Test parameters		
Year	2000	
Subject	Mathematics, reading, science	
Type of school	Public, private	
Grade	15-year-olds (mainly 9th and 10th graders)	
Number of students tested	5,190	
Test costs	**Soles**	**US$ (PPP)**
Cost of participation	150,000	100,000
Test preparation	257,000	171,000
Test application	193,000	128,000
Processing and analysis	73,000	49,000
Dissemination	49,000	33,000
TOTAL	**722,000**	**480,000**
Cost per student	139	92
Cost of educating a student[a]	832	553
Cost of testing as % of total expenditures on secondary education	0.04	

SOURCE: UMC.

NOTES: 2000 PPP exchange rate: 1.5 soles = US$1. Costs of UMC staff were not included in test preparation and processing and analysis.

[a]Cost per student at the secondary level in 2000. See Wolff and Gurría (2005).

Preparation of the test cost $171,000. This included designing the sample and test items, traveling to PISA meetings around the world, and reviewing Spanish translations. Peru's higher costs relative to Uruguay are in part due to the difficulties of undertaking a nationally representative sample in a much larger and mountainous country. Peru spent only $33,000 on dissemination. In any event, the total cost for PISA participation was $480,000, which represented only 0.04 percent of Peru's annual expenditures on secondary education. Breakdown of costs by activity was as follows: PISA membership, 21 percent; preparation, 36 percent; application, 27 percent; processing and analysis, 10 percent; and dissemination, 7 percent.

Colombia

The costs of Colombia's participation in PIRLS to test reading competency in the fourth grade in 2001 are similar to those of Peru for PISA participation, as can be seen in table 10—notwithstanding the fact that Colombia did not report its costs for analysis and dissemination. Its reported costs of test application were higher than Peru's.

Table 10: Colombia—Participation in PIRLS

Test parameters		
Year	2001	
Subject	Reading	
Type of school	Public, private	
Grade	4	
Number of students tested	5,131	
Test costs	**Pesos**	**US$ (PPP)**
Cost of participation	41,220,000	60,000
Test preparation	93,986,000	137,000
Test application	276,582,000	402,000
Processing and analysis	NA	NA
Dissemination	NA	NA
TOTAL	**411,788,000**	**599,000**
Cost per student	80,255	117
Cost of educating a student[a]	840,228	1,223
Cost of testing as % of total expenditures on primary education	0.01	

SOURCE: ICFES.

NOTES: NA = not available. Overhead costs are not included. 2001 PPP exchange rate: 688 pesos = $US1.

[a]Cost per student at the primary level in 2000. See Wolff and Gurría (2005).

Costs of Participating in UNESCO's Regional Test

Costs for participating in UNESCO's regional testing program, LLECE, are also available. This program tested third and fourth graders in 1997 and 1998 in language and mathematics in 13 countries. Table 11 provides an estimate of costs for central management as well as for the 13 participating countries.

Table 11: Estimated regional costs of LLECE test, 1997–99

Funding sources	US$
Direct support from countries	269,000
UNESCO support	80,000
Ford Foundation	50,000
IDB	500,000
Subtotal	**899,000**
In-country costs[a]	1,100,000
TOTAL	**1,999,000**

SOURCE: IDB estimates based on documentation provided by UNESCO.

[a]Costs are estimated for 12 countries.

The total cost for the entire program was just under $2 million, or about $154,000 per country—far lower than participating in PISA or PIRLS. These lower costs are a result, in part, of the fact that each country had to contribute only $10,000 per year to UNESCO, compared to $30,000 to $50,000 per year to participate in PISA, TIMSS, or PIRLS. Travel for expert meetings was much less costly for a regional study than for an international study. On the LLECE test, all but one of the items were multiple choice. Sample size was around 5,000 students in each country. This round of testing was criticized for having a number of technical inconsistencies, which may have been a result, in part, of inadequate funding of specialists. The costs for the current round of testing by LLECE are expected to be significantly higher than the first round. LLECE is keeping its costs down by using Latin American experts within country testing institutions to undertake much of the preparatory work; it has established an international oversight committee to ensure technical quality. A preliminary estimate of the total costs, as reported by LLECE staff, is about $5.2 million for 17 countries and one state, or around $307,000 per participant.

III. Conclusions about the Costs of Testing in Latin America

Table 12 and figures 1, 2, and 3 summarize the information collected on the costs of testing. This section highlights some conclusions that can be drawn about the costs of testing in Latin America.

Testing Costs as a Percentage of Education Costs

As depicted in figure 1, testing as currently practiced in the region is not a significant financial burden—constituting well below 1 percent of the total budget of the level of education (primary or secondary) tested.

The costs of participation in international programs, which usually require samples of around 5,000, range in the countries studied from $300,000 to $600,000. In Colombia, testing is 0.02 percent or less of the national budget of primary or secondary education. In most other countries, the costs of testing varies between 0.02 percent and 0.17 percent of the national budget. Honduras, probably because of its need to rely on external consultants and because of its decision to outsource all aspects of the testing process, shows the highest relative cost, which is still only 0.3 percent of the national primary public schooling budget. Of course, more expenditures are necessary to implement a test of all students in a grade rather than a sample, as is discussed further below. For example, Colombia spent $2.5 million to test all students in one

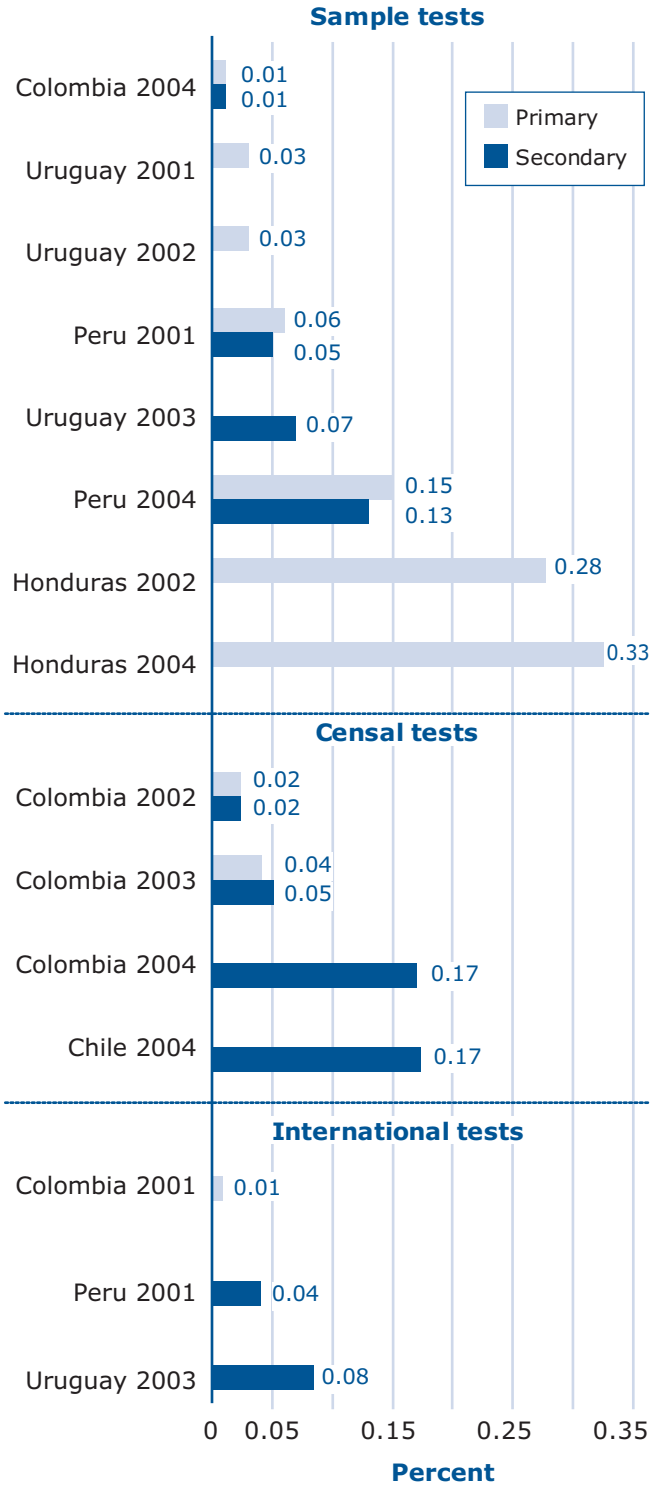
Table 12: Summary of costs of testing in Latin America

Country	Type of test	Year	Subject	Grade	Total cost (US$ PPP)	Number of students tested	Cost per student (US$ PPP)	Cost of testing as % of total budget for educational level tested[a]	
								Primary	Secondary
Chile	Censal	2004	Language, mathematics, natural science, social science	8	4,472,000[b]	300,000	15		0.17
Colombia	Int'l	2001	Reading	4	599,000	5,131	117	0.01	
	Censal	2002	Language, mathematics	5, 9	2,548,000	1,030,626	2	0.02	0.02
	Censal	2003	Citizenship, science	5, 9	5,497,000	1,034,049	5	0.04	0.05
	Censal	2004	Higher education entrance exam	11	9,418,000	540,716	17		0.17
	Sample	2004	Language, mathematics	5, 9	1,363,000	96,242	14	0.01	0.01
Honduras	Sample	2002	Language, mathematics, natural science	3, 6	1,953,000	42,572	46	0.28	
	Sample	2004	Language, mathematics, natural science	3, 6	2,292,000	45,657	50	0.33	
Peru	Int'l	2000	Mathematics, reading, science	15-year-olds in grade 7 or higher	480,000	5,190	92		0.04
	Sample	2001	Mathematics, reading and writing	4, 6, 10	1,757,000	34,000	52	0.06	0.05
	Sample	2004	Citizenship, mathematics, reading and writing	2, 6, 9, 11	4,885,000	70,000	70	0.15	0.13
Uruguay	Int'l	2003	Mathematics, problem solving, reading, science	15-year-olds in grade 7 or higher	311,000	5,797	54		0.08
	Sample	2001	Cognitive and affective development, language, mathematics	Preschool, 1, 2	102,000	2,387	43	0.03	
	Sample	2002	Language, mathematics	6	111,000	9,171	12	0.03	
	Sample	2003	Language, mathematics, natural science, social science	12	266,000	12,993	20		0.07

[a]The total budget for education by level corresponds to their respective values in 2000. When a test is administered in primary as well as secondary education it is assumed that each level bears half of the costs of testing. Overhead costs are not included. For possible explanations for some of the absolute total and unit cost differences found in this table, see section "Variability of Testing Costs" below.

[b]Total does not include overhead costs.

Figure 1: Testing costs as a % of the budget of the corresponding level of study



grade but about $1.2 million to test a large sample in 2004. Note too that no country in the region tests all students in all grades.

While none of these are extraordinarily large amounts, it must be remembered that educational and public sector budgets in Latin America are fairly small, due to low gross domestic product (GDP) and low tax revenues—not to mention unclear priorities. Teacher and administrative personnel salaries tend to absorb a lion's share of those resources—between 70 and 91 percent of total expenditures (Bruneforth, Motivans, and Zhang 2004), and "discretionary" budgets are fairly limited, a situation not easily resolved in the short term. It is likely that testing will be perceived as a greater burden for small, poor countries (e.g., those with a per capita GDP below $5,000) with limited budgetary and human resource capacity, as seems to be the case in Honduras, as opposed to the larger, more economically advanced, countries.

On the other hand, testing seems to be inexpensive compared to most interventions proposed to improve learning. A recent estimate of costs of primary education gives an idea of the costs of testing compared with other interventions commonly discussed and implemented in Latin America (see table 13).

Testing is among the least expensive innovations in primary education reform, requiring as little as 1 percent of the cost of increasing teachers' salaries, reducing class size, and/or reforming teacher training. This suggests that testing can be a relatively inexpensive supplement to reform efforts—provided that results are technically adequate and used for decision making. Of course, tests whose technical quality is so low as to preclude drawing any valid conclusions about learning, or where the information is neither disseminated nor used, are very expensive indeed. At the same time, there are hidden costs in testing that have

Table 13: Estimated costs of potential interventions in primary education

Type of intervention	Estimated increase in unit costs (%)
Test sample of fourth graders	0.01–0.3
Test all fourth graders	0.04–0.2
Reduce class size by 10%	9
Preschooling for all at-risk children (50%)	8.3
Raise teacher salaries by 20%	18
Targeted in-service training of teachers (one week per year)	2.3
Interactive radio instruction	0.5

SOURCE: Adapted from Schiefelbein, Wolff, and Schiefelbein (2000).

not been taken into account here. In addition to better estimates of overhead costs, a full cost accounting should take into consideration the time that teachers might, in some countries, take to administer tests, collect materials, and provide test results to supervisors, which could easily be the equivalent of an entire day of schooling. If all the students in a particular grade were tested, this would be about 0.55 percent of the total costs for that grade, assuming 180 days in the school year—greatly exceeding reported testing costs. It would also mean less time for teaching. In addition, school supervisors are likely to spend time preparing and implementing tests for which they are not directly reimbursed. And, as noted above, there are additional costs associated with changing curriculum, training teachers, and other programs that derive—or should derive—from testing.

Variability of Testing Costs

The costs of testing vary greatly from one country to another, even when the countries are undertaking similar programs. For example, Colombia tested 96,000 students at a cost of $1.4 million, while Honduras's cost for testing 46,000 students was $2.3 million and Peru's for 70,000 students was $4.9 million. Participating in PISA cost $480,000 in Peru and $311,000 in Uruguay; and Colombia's participation in PIRLS cost $599,000. The differences in total costs between countries can be explained in part by differences in the number of grades and subjects tested. Figure 2 shows (1) the total costs of testing for each country in a particular year and for a particular type of test (e.g., census, sample, or participation in an international test) and (2) what the cost of testing only one grade and one subject would be (adjusted cost).

Costs per student tested in national samples and censal applications also vary significantly, from Colombia's cost of about $2 per student to Peru's $70. As with total costs, however, relative costs vary when the number of grades and subjects being tested are considered, as shown in figure 3.

Thus, some possible explanations for absolute total and unit cost differences from one country to another include the following:

- Testing multiple subjects and grades is more costly than testing one or two subjects and one grade. Table 14 shows an estimate of costs on the basis of one subject per grade. With this approach, testing costs in Peru decline significantly in comparison with the other countries in the study, as was shown in figure 3. For example, Peru tested four grades and three subjects in 2004, for a total of 12 tests, compared with 4 or 6 for most other countries. Similarly, in 2001, Uruguay tested four subjects and three grades, also for a total of 12 tests.
- Countries with large populations and land area are more costly to sample than small, compact countries.
- Open-ended questions are more costly to score than multiple-choice questions. Colombia's exclu-

Figure 2: Total and adjusted costs of testing

NOTE: The adjusted cost was obtained by dividing the total cost by the number of grades and subjects tested.

Figure 3: Cost and adjusted cost per student tested

NOTE: The adjusted cost per student was obtained by dividing the cost per student by the number of grades and subjects tested.

Table 14: Costs of testing by subject/grade

Country, year, and type of test	Total unit cost (US$)	Number of grades tested	Number of subjects tested	Total number of grades and subjects tested	Unit cost per grade/subject tested (US$)
Chile 2004 C	15	1	4	4	3.75
Colombia 2001 I	117	1	1	1	117.00
Colombia 2002 C	2.5	2	2	4	0.63
Colombia 2003 C	5.3	2	2	4	1.25
Colombia 2004 S	14	2	2	4	3.50
Honduras 2002 S	46	2	3	6	7.70
Honduras 2004 S	50	2	3	6	8.30
Peru 2001 I	92	1	3	3	30.70
Peru 2001 S	52	3	2	6	8.67
Peru 2004 S	70	4	3	12	5.80
Uruguay 2003 I	54	1	4	4	13.50
Uruguay 2001 S	43	3	4	12	3.60
Uruguay 2002 S	12	1	2	2	6.00
Uruguay 2003 S	20	1	4	4	5.00

Note: C = censal; S = sample; I = international.

sive use of multiple-choice questions is a factor in its relatively lower costs, whereas Peru's decision to explore the processes by which students arrive at a given response to a test item via the use of nearly 50 percent open-ended questions certainly contributed to the latter country's higher testing costs.

- Countries with limited human resources may need to draw on—and pay more for—local and foreign consultants, compared to countries with greater local human resource capacity.
- Careful and widespread pilot testing, security control, analysis, and dissemination increase costs.
- Effective computerization and strong management can reduce costs, as was reported by Colombian authorities.
- Per student costs of testing small samples are much higher than costs of testing larger samples or all students.
- Support from international lending or grant agencies may lead to increased costs because of procurement regulations, as well as simply because of the availability of these funds.
- There is uncertainty on how to account for overhead costs of public agencies administering a country's testing program. In this report, overhead costs have not been consistently and fully considered. Costs could increase by as much as 60 percent, as is the case in one of the countries studied, if overhead expenses are included. Chile calculated its overhead costs but was not able to allocate these to the various tests it was administering. Colombia did not estimate its overhead costs, which are likely to be significant. Peru estimated its costs with and without overhead; the former were significant. Honduras did not have government overhead costs since its testing unit was located in a university under contract with the government. Uruguay did not estimate overhead costs.

- Wage scales may vary from one country to another even after correcting for PPP.
- Staffing issues can have an impact on financing. Most countries in the region have only a very small cadre of experts in testing, and they are often overcommitted. Inadequate staffing often leads to higher costs because of the need to contract for expensive external consultants, and/or problems in technical quality and dissemination. Over the long run, training more local experts will not only lead to better technical quality but also to lower costs.

In short, tracking testing costs is complex, given uncertainties in what is defined as overhead and/or dissemination, in considering overhead and hidden costs, and in trying to make costs comparable from one country to another.

IV. Costs and Decision Making about Testing

Testing—at least on a sample basis—is a fundamental element in any modern education system, since it is a necessary step in the process of designing, implementing, and evaluating reform efforts.[8] Beyond this generalization, decision making about the kinds and scope of tests to be undertaken depends on a wide variety of factors, starting with the definition of feasible learning and other objectives and leading to considerations of technical and management capacity. Costs play an important, but not defining, role in these decisions. Each country has a different set of conditions, and decision makers and technicians need to make their own trade-offs in breadth and depth of testing based on their objectives. Below are a few suggestions that ought to be taken into consideration as country officials ponder what kinds of tests to administer.

[8]Uruguay is the best example in the region of a country that has used testing programs as a baseline to measure whether a major reform has had an impact on learning, and with positive results.

Difficulty of Measuring Results in Education

Countries must recognize that measuring results in education through testing is not a simple matter. A recent study (Crone 2004) tried to summarize major caveats regarding testing. To begin with, a test must accurately reflect what is proposed to be taught. This requires careful item development, pretesting, and analysis. Second, an individual score will vary each time a student takes a test not only because of particular test items included but the physical or emotional status of the student as well. Third, it is difficult to determine through testing whether a particular teacher, school, or system is more effective than another. The fundamental problem is that the raw material, the student, is not homogeneous—children and youths have various backgrounds, capacities, and experiences. Therefore, a teacher whose students score only at the mean in a standardized test, but whose students' parents are illiterate, could be considered more effective than one whose students score in the top 25 percent but whose parents are rich and college educated. It is also difficult to measure the value added of learning from one year to another unless there are unique student identification numbers. Some of the increases in test scores do not reflect increased learning per se but instead reflect increased experience and comfort with the test-taking process itself. Finally, states or countries can lower their standards. A number of U.S. states have reported improvements in scores on statewide tests, while at the same time, the National Assessment of Educational Progress has shown little or no progress in these same states. There is a suspicion that these states are lowering testing standards so as to report increased learning (Dillon 2005).

Trade-Offs in Testing Goals, Complexity, Capacity, and Cost

As noted above, decision making about testing starts with identifying and selecting relevant learning and other goals, identifying and procuring endorsement and support or interest of key stakeholders, and recognizing constraints in terms of human resources and costs. Taking all these issues together, decision makers will need to make and accept trade-offs on a wide variety of parameters.

The most important rule of thumb in this process is never to be sparing with technical quality or dissemination, since a technically inadequate test or one that is not used or disseminated is a real and unacceptable waste of money. Testing objectives must be kept simple and clear.

When in doubt, it might be better to test fewer grades and subjects and concentrate on dissemination of results. In fact, testing many grades and subjects at one time, while saving on staff, materials, and transportation costs, may overwhelm the ability of the testing agency to analyze, report and—most importantly—disseminate results effectively; this may well have been the case in Peru. On the other hand, when subjects are *not* tested, teachers and students may believe that they are not important.

Another critical issue to be dealt with is whether to undertake a sample or census. As illustrated in table 1, this decision should be made on the basis of a clear definition of objectives and form of utilization of the results. Sample size also has a strong impact on the cost per student tested. In Colombia, for example, a sample of around 5,000, for participation in international tests costs $117 per student. In a larger sample (96,000 students), the cost per student tested is $14; and in censal testing, the cost is between $2 and $5.

Fixed and variable costs need to be taken into account. The fixed costs of testing are those of preparing and validating items, including pilot testing in the field. The costs of preparing items can be reduced if only a small portion is made available to the public, therefore permitting reuse of already validated items. The variable costs are those of application in the field and dissemination of results. Analysis of results is a mix of fixed and variable costs. Powerful computers now permit analysis of large data sets at little additional cost, but correcting open-ended items is a variable cost. Correcting open-ended questions can also be time consuming and costly, even though open-ended items are usually more effective measures of higher order skills. When Colombia did a censal test, its fixed costs for test preparation were equal to only 6.6 percent of the total costs. When it did a sample, its fixed costs were equal to about 10.6 percent of total costs. In PIRLS, with less than 6,000 students, its test preparation costs were 25.4 percent of total costs. Small samples can be cost effective, but, at the same time, large samples can be expanded to censal testing at a low marginal cost, since the fixed costs of developing items and pilot testing can be amortized over a larger population.

Taking into account the current level of capacity, testing all children in every grade every year—as is done in many states in the United States—does not appear to be a good idea for countries in Latin America. Chile probably has the most extensive censal testing program, but, until recently, it had been testing no more than one grade a year for grades 1–8. Implementing full annual testing in Chile would increase expenditures close to eightfold, leading to costs that would be over 1 percent of national educational expenditures, even assuming some economies of scale. The proportional cost would be even higher in Peru and Honduras. In comparison, the cost in the United States is, on average, about 0.25 percent of total per

student annual education costs. Of course, this lower ratio derives from the much higher expenditures for teaching and teachers in the United States.

Certainly, if spending 1 percent of education costs on testing could result in major increases in student achievement, then the result might be worth the expense. But massive testing would likely overwhelm the more important task of disseminating test results and designing teacher training and other interventions based on test results, without which there will be no improvement, and which will demand considerable additional public expenditures. Testing only some grades and subjects will reveal progress and challenges for schools, communities, provinces, and the country as a whole and will help parents to place their children in a larger context. Thus, testing one or two grades per year achieves nearly all of the goals of education testing except feedback to all parents. Testing everyone in every grade every year increases the negative aspects of this process, such as taking time from teaching; neglecting less frequently tested subjects such as history, art, and science; and discouraging innovation in the classroom.

Emphasizing Dissemination

Without adequate dissemination of results, testing will have no impact on learning or achievement, since the results will not be used by teachers, parents, school principals, and decision makers. Therefore, expenditures on dissemination should never be skimped, and a dissemination plan should be in place before beginning any testing program. As currently practiced in Latin America, dissemination is not costly compared to the rest of the testing process. Reported dissemination costs, usually for publication of results, ranged between 5 and 15 percent of total testing costs. Fully costing dissemination should also include analysis by decision makers and changes in teacher training, curriculum, and textbooks. For a detailed discussion of the current state of dissemination of testing results in Latin America, see Ravela (2002) and Cueto (2005).

Potential Impact of High-Stakes Testing

High-stakes tests at the secondary level, or curriculum-based external exit exams (CBEEEs), determine whether a student graduates from a particular level of education and can affect his/her options in higher education. Such CBEEEs are common to countries as varied as Britain, France, Denmark, Japan, South Korea, the Netherlands, and Germany. They differ from college entrance examinations in that they are universal, necessary for graduation, and curriculum based. There is increasing evidence around the world that high-stakes testing designed to certify that students have successfully completed secondary education can, in itself, increase the level of learning achievement. Two studies by Bishop (1997 and 2003) and another by Woessman (2000) concluded that high-stakes testing increased learning at the secondary level. Of the 40 countries involved in TIMSS, those with national high-stakes testing tended to score higher than those without such high-stakes tests.[9] Within the United States, states with CBEEEs, such as New York (the Regents exam at the end of high school), have better student achievement results than states that do not have high-stakes exams. These results are independent of per student expenditures.

Several developments in the United States build on these conclusions. For example, the Scholastic Aptitude Test (SAT) is becoming increasingly based on actual achievement rather than on "generalized" learn-

[9]The Bishop study also concluded that another reason for the low performance of U.S. students is that their teachers are grossly underpaid and overworked as compared to teachers in Europe and Japan.

ing skills, which is a way of increasing the "stakes" of high school learning. Furthermore, use in the United States of the Advanced Placement exams of the Educational Testing Service, which are also high-stakes exams for high school students seeking entrance to elite colleges and universities, is becoming increasingly widespread.

All English-speaking Caribbean countries have high-stakes exams, but most of the Latin American countries in the region do not. The exceptions are Costa Rica and the Dominican Republic, where a test given at the end of secondary education counts for at least 25 percent of a student's final grade; and El Salvador, where the Learning and Aptitude Test for High School Students (Prueba de Aptitudes para Egresados de Educación Media) accounts for 20 percent of the final grade. There is as yet no evidence that students in these countries perform better than expected because of these exams. Chile has begun to consider high-stakes testing at the end of secondary education, and the Peruvian National Council of Education has proposed that serious consideration be given to this option. This report does not provide data on the costs of these tests, except for ICFES. The tests are likely to be relatively expensive, but the payoff in terms of learning achievement can be significant.

High-stakes testing has obvious risks. Care must be taken that tests are designed to emphasize higher order learning rather than memorization, and to ensure that students do not drop out in advance of taking a high-stakes exam. This last was the case in Texas a few years ago, although dropout rates are now reported to be stabilized (Carnoy and Loeb 2003). Dropout in advance of testing has also been anecdotally reported in the English-speaking Caribbean.

A variation of high-stakes testing used in the United States is "minimum competency" exams, where all students must achieve a "floor" of learning in order to graduate. The evidence to date (see Bishop 1997) is that minimum competency exams do not have an observable impact on learning and may lead to increases in the dropout rate.

Value of Participating in International Testing Programs

Participation in international testing can launch a national debate and reform efforts, which can in turn lead to increased learning and create a much-valued forum for national assessment capacity development. Around the world, low scores on international tests have led to such debates and a wide range of programs to improve learning in the various participating countries. For example, Germany scored much lower than expected on PISA and is now rethinking its policy of early streaming of many students into low-expectation programs (Ammermueller 2004). Some places (South Korea, Hong Kong, the United States, and Ontario, Canada, among others) have recently improved learning, as measured by scores on the TIMSS examinations of science and mathematics between 1995 and 2003. The reasons for these improvements require analysis beyond the scope of this paper. Within Latin America, the poor results of Chile, Mexico, Peru, and Brazil on both TIMSS and PISA have similarly triggered national debates, although positive outcomes of such discussions largely remain to be seen. It is quite possible that Latin American students do poorly in international tests because the opportunity to learn is lacking—e.g., what is measured in the test is not taught in the classroom. For example, while many international tests for language ask students to write essays, or for mathematics, to solve word problems, these skills reportedly are rarely taught in the typical Latin American classroom.

While participation in an international testing program is not a significant financial burden, and is

much valued as a training ground by almost all Latin American participants, other problems frequently make it difficult for some countries to participate and must be taken into account by decision makers. For example, PISA and PIRLS meetings can be held anywhere in the world. Long-distance travel to, say, the Far East, often has to be cleared with top-level government officials, especially in smaller countries, and can be difficult when countries are implementing fiscal austerity measures. Education decision makers need to convince budget officials of the importance of these international initiatives.

In addition, participating in international programs can make for very high demands being placed on a small cadre of qualified in-country experts, who are usually simultaneously involved in national testing efforts. The long-term solution is to train more specialists in the area of testing.

Participation in the regional UNESCO tests seems to be more attractive to some countries—particularly in terms of viability of closer involvement, learning by doing, and lesser costs. The UNESCO tests could help achieve many of the goals of international tests if portions of the tests could be made equivalent to PISA, PIRLS, etc.

Importance of Identifying What Works in Education

Standardized testing is not expensive compared to total education expenditures in Latin America. A more fundamental question is to identify what works in education. Unfortunately, education systems throughout the world are notorious for under-investing in research and development, and Latin America is no exception.

As noted above, research suggests that high-stakes testing can increase learning. There is evidence, albeit inconclusive, that feedback of test results that then lead to rewards/punishments for teachers and schools (i.e., accountability) for increased learning will positively affect learning. In the United States, the No Child Left Behind Act mandates action to improve test scores and includes a variety of rewards and punishments to schools and communities for increased scores. A recent study (Carnoy and Loeb 2003) concluded that, while promising, it was still too early to identify clearly whether teacher, school, and community accountability in the United States has had an impact on learning. Within Latin America, Chile has targeted poorly performing districts (the 900 Schools Program) for additional assistance. In the past, there has been improvement in test scores in these schools, but there was also evidence that some schools did not want to "graduate" to better performing levels since they would then lose the extra resources accorded them. Chile's system of encouraging teacher quality (the National System of School Performance Assessment—Sistema Nacional de Evaluación de Desempeños de los Establecimientos Educacionales) is well accepted by teachers, but it has not yet been demonstrated to have an impact on learning. Nor is there evidence that Mexico's incentives to teachers for higher student scores on tests have improved learning performance (see Mizala and Romaguera 2005 for a study of both countries' experiences). Feedback of test results to reform curriculum, revise textbooks, and improve teacher training ought to have an impact on learning, but there is little direct evidence of the impact, largely because of the complexity of the process. In short, much remains to be done to measure the impact of education policies and programs on learning.

Bibliography

Administración Nacional de la Educación Pública. 1998. *Evaluación de aprendizajes en tercer año de educación primaria.* Montevideo, Uruguay.

Ammermuller, Andreas. 2004. *PISA: What Makes the Difference? Explaining the Gap in PISA Test Scores between Finland and Germany.* Discussion Paper No. 04-04. Munich: Center for European Economic Research.

Bishop, John. 1997. "The Effect of National Standards and Curriculum-Based External Exams on Student Achievement." *American Economic Review* 87(2):260–64.

———. 2003. "La Educación Secundaria en los Estados Unidos. ¿Qué pueden aprender otros de nuestros errores?" In G. Rama, ed., *Alternativas de Reforma de la Educación Secundaria*, pp. 63–108. Washington, DC: Inter-American Development Bank.

Bruneforth, Michael, Albert Motivans, and Yanhong Zhang. 2004. *Investing in the Future: Financing the Expansion of Educational Opportunity in Latin America and the Caribbean.* Montreal: UNESCO Institute for Statistics.

California Legislative Analyst's Office. 2003. "Analysis of the 2003–2004 Budget Bill, Education Table of Contents." Sacramento.

Carnoy, Martin, and Susanna Loeb. 2003. "Does Accountability Affect Student Outcomes? A Cross State Analysis." *Education Evaluation and Policy Analysis* 24(4).

———. 2004. *¿Tiene efecto la responsabilidad externa en los resultados educacionales de los alumnos?. Un análisis entre los estados de EE.UU.* Working Paper No. 26. Santiago: PREAL.

Cohen, Joel, and David Bloom. 2005. "Cultivating Minds." *Finance and Development* 42(2).

Connecticut State Department of Education. 2005. "Cost of Implementing the Federal No Child Left Behind Act in Connecticut." Hartford.

Crone, Theodore. 2004. "What Test Scores Can and Cannot Tell Us about the Quality of Our Schools." *Business Review* Q3:5–21.

Cueto, Santiago, ed. 2005. *Uso e impacto de la información educativa en América Latina.* Santiago: PREAL.

Dillon, Sam. 2005. "Students Ace State Tests, but Earn D's from US." *New York Times* November 26.

Ferrer, Guillermo. 2006. *Educational Assessment Systems in Latin America: Current Practice and Future Challenges.* Washington, DC: PREAL.

Ministerio de Educación de Chile. 2004. *Resultados nacionales de Octavo Básico.* Santiago.

Mizala, Alejandra, and Pilar Romaguera. 2005. "Rendimiento escolar y premios por desempeño: la experiencia latinoamericana y el SNED en Chile." In Santiago Cueto, ed., *Uso e impacto de la información educativa en América Latina,* pp. 23–60. Santiago: PREAL.

Peyser, James, and Robert Costrell. 2004. "Exploring the Costs of Accountability." *Education Next* Spring:23–29.

Ravela, Pedro. 2002. *¿Como Presentan sus Resultados los Sistemas Nacionales de Evaluación Educativa en América Latina?* Working Paper No. 22. Santiago: PREAL.

Ravela, Pedro, ed. 2001. *Los Próximos Pasos: ¿Cómo avanzar en la evaluación de aprendizajes en América Latina?* Working Paper No. 20. Santiago: PREAL.

Schiefelbein, Ernesto, Laurence Wolff, and Paulina Schiefelbein. 2000. "La opinión de expertos como instrumento para evaluar la inversión en educación primaria." *Revista de la CEPAL* 72:147–63.

Woessmann, Ludger. 2000. *Schooling Resources, Educational Institutions, and Student Performance: the International Evidence.* Working Paper No. 983. Kiel, Germany: Kiel Institute of World Economics.

Wolff, Laurence, and Martin Gurría. 2005. "Money Counts: Projecting Education Expenditures in Latin America and the Caribbean to the Year 2015." UIS Working Paper.

Web Sites

Chile: National System for the Assessment of Educational Quality (Sistema de Medición de la Calidad de la Educación—SIMCE), www.simce.cl.

Colombia: Institute for the Development of Higher Education (Instituto Colombiano de Fomento de la Educación Superior—ICFES), www.icfes.gov.co.

Honduras: Unit for Measuring Educational Quality (Unidad de Medición de la Calidad Educativa—UMCE), www.se.gob.hn.

Peru: Quality Measuring Unit (Unidad de Medición de la Calidad Educativa—UMC), www.minedu.gob.pe/umc.

Uruguay: National Administration for Public Education (Administración Nacional de la Educación Pública—ANEP), www.anep.edu.uy.

Latin American Laboratory for the Measurement of Quality in Education (Laboratorio Latinoamericano de Medición de la Calidad de la Educación), www.llece.unesco.cl.

Appendix: Summary of National Educational Assessment Systems

The following text and tables are excerpted from Ferrer (2006) and include some basic information about the characteristics of the national educational assessment systems of the Latin American countries discussed here—Chile, Colombia, Honduras, Peru, and Uruguay—that should be taken into account when comparing costs.

Chile

All tests are census-based. Traditionally, they have been norm-referenced, although recently the model has been redefined in such a way that the results report can be criterion-referenced. The instruments include multiple-choice items as well as open-ended questions. All assessment exercises are complemented by context questionnaires for later analysis of in-school and out-of-school factors related to academic performance. Additionally, a special technical team quantitatively and qualitatively monitors the implementation of the curriculum at the national level by

Table A-1: Chile—Summary of national educational assessment system

Test year	Grades tested	Subjects tested	Test coverage	Results reported to	Results used for
Domestic assessments: Prueba de Evaluación de Rendimiento Escolar (School Achievement Test)					
1982–1984	4, 8	■ Language ■ Mathematics ■ Natural sciences ■ Social sciences	■ National census ■ National sample	■ Government ■ Users ■ Public	—
Domestic assessments, administered by SIMCE					
1988	4	■ Language ■ Mathematics ■ Natural sciences ■ Social sciences ■ Student attitudes	■ National census ■ Experimental sample	■ Government ■ Users ■ Public	—
1989	8				
1990	4				
1991	8				■ Teacher training ■ Curricular development ■ Targeting support to students and schools
1992	4				
1993	2M	Student attitudes			
1994	4, 8	■ Language ■ Mathematics ■ Natural sciences ■ Social sciences ■ Student attitudes			
1995	2M	■ Language ■ Mathematics ■ Student attitudes			
	8	■ Language ■ Mathematics ■ Natural sciences ■ Social sciences			
1996	4	■ Language ■ Mathematics ■ Natural sciences ■ Social sciences ■ Student attitudes			
1997	8				

surveying and observing teachers and principals. This monitoring facilitates more in-depth research on the impact different sources of curricular design used in the schools (curricular frameworks, syllabuses, textbooks, etc.) have on students' academic attainment.

Table A-1: Chile—Summary of national educational assessment system (continued)

Test year	Grades tested	Subjects tested	Test coverage	Results reported to	Results used for
Domestic assessments, administered by SIMCE					
1998	2M	■ Language ■ Mathematics	■ National sample ■ Experimental sample	■ Government ■ Users ■ Public	■ Teacher training ■ Curricular development ■ Targeting support to students and schools
1999	4	■ Language ■ Mathematics ■ Natural sciences ■ Social sciences	■ National census ■ Experimental sample		
2000	8				
2001	2M	■ Language ■ Mathematics			
2002	4	■ Language ■ Mathematics ■ Natural sciences ■ Social sciences			
2003	2M	■ Language ■ Mathematics			
2004	8	■ Language ■ Mathematics ■ Natural sciences ■ Social sciences	National census		■ Teacher training ■ Curricular development ■ Targeting support to students and schools ■ Teacher incentives
International assessments					
■ Six Subject Study (1970–71) ■ LLECE (1997) ■ TIMSS-R (1998) ■ International Adult Literacy Survey (1998)			■ IEA-Civic Education (2000) ■ PISA Plus (2001) ■ TIMSS (2003)		

Colombia

The SABER tests are sample-based and criterion-referenced, and they assess academic attainment in accordance with a predefined performance scale structured on three levels of increasing complexity for each subject and year being assessed. The tests are complemented by context questionnaires that facilitate analysis of the in-school and out-of-school factors related to academic performance (characteristics of the school, teacher, student, and family). The state examinations are administered on a census basis to all those seeking admission to higher education. They assess breadth of knowledge in curricular areas of general learning and facilitate assessment of in-depth knowledge in certain disciplines, in line with the admission requirements of various higher education programs. Also assessed are foreign language skills and interdisciplinary learning (specifically, the environment and violence and society).

Table A-2: Colombia—Summary of national educational assessment system

Test year	Grades tested	Subjects tested	Test coverage	Results reported to	Results used for
Domestic assessments, administered by ICFES/Ministry of Education					
1991–1994	3, 5, 7, 9	■ Language ■ Mathematics	■ National sample ■ Regional sample	■ Government ■ Users ■ Public	—
Domestic assessments: SABER tests					
1997–1998	3, 5, 7, 9	■ Language ■ Mathematics	National sample	■ Government ■ Users ■ Public	—
1998–1999	7, 9	■ Language ■ Mathematics ■ Natural sciences			
2002–2003	5, 9[a]		National census	—	
Domestic assessments: ICFES state examinations					
1980–2005	11	■ Foreign language ■ Language ■ Mathematics ■ Natural sciences ■ Social sciences	National census	■ Government ■ Users	■ Curricular development ■ Selection for higher education
International assessments					
■ TIMSS (1995) ■ LLECE (1997)			■ IEA-Civic Education (2000) ■ PIRLS (2001)		

[a]In some areas, grades 3 and 7 were also tested.

Honduras

UMCE's tests are sample-based at the national level, criterion-referenced, and based mainly on a multiple-choice model. Both written composition and oral reading are tested on the verbal assessments. The tests are complemented by questionnaires for students, parents, teachers, and principals to gather data for later analysis of school-related factors. Attention is focused on socioeconomic status and other intra-school variables (including identification of a model for gauging the effectiveness of schooling). The technical teams have faced some difficulties in establishing reliable correlations between these variables and academic performance, although the use of a hierarchical linear model since 2002 has enhanced the reliability and validity of the correlations. UMCE additionally offers its services in assessing academic performance as an indicator of the impact of specific programs financed by international development institutions.

Table A-3: Honduras—Summary of national educational assessment system

<table>
<tr><th>Test year</th><th>Grades tested</th><th>Subjects tested</th><th>Test coverage</th><th>Results reported to</th><th>Results used for</th></tr>
<tr><td colspan="6">Domestic assessments, administered by Ministry of Education[a]</td></tr>
<tr><td>1990–1994</td><td>1, 2, 3, 4, 5</td><td>■ Language
■ Mathematics
■ Natural sciences
■ Social sciences</td><td>Experimental sample (10 municipalities)</td><td>Education Secretariat
USAID</td><td>Assessing USAID-supported projects</td></tr>
<tr><td colspan="6">Domestic assessments, administered by UMCE</td></tr>
<tr><td>1997</td><td>3, 6</td><td rowspan="3">■ Language
■ Mathematics</td><td>■ National sample
■ National census</td><td rowspan="4">■ Government
■ Users</td><td>■ Teacher training
■ Targeting support to students and schools</td></tr>
<tr><td>1998</td><td>2, 3, 4, 6</td><td rowspan="3">National sample</td><td rowspan="3">—</td></tr>
<tr><td>1999</td><td>2, 3, 4, 5</td></tr>
<tr><td>2000</td><td>3, 6</td></tr>
<tr><td>2002–2004</td><td>3, 6</td><td>■ Language
■ Mathematics
■ Natural sciences</td><td></td></tr>
<tr><td colspan="6">International assessments</td></tr>
<tr><td colspan="6">LLECE (1997)</td></tr>
</table>

[a]Assessment was undertaken as Component V of the Primary Education Enhancement Project.

Peru

Peru's first tests were norm-referenced, but they are now criterion-referenced. Tests are sample-based; include both multiple-choice and open-ended questions; and have been administered in Spanish, Quechua, and Aymara. The tests are complemented by context questionnaires for later analysis of performance-related factors. These questionnaires focus on both in-school variables (teaching inputs, characteristics of schools and teachers, attitudes toward subjects and indigenous languages, public or private school management) and out-of-school variables (gender, socioeconomic status, native language, household chores, urban or rural location, geographic region).

Table A-4: Peru—Summary of national educational assessment system

Test year	Grades tested	Subjects tested	Test coverage	Results reported to	Results used for
Domestic assessments, administered by UMC					
1996	4	■ Language ■ Mathematics	National sample	—	—
1998	4, 6, 4M, 5M	■ Language ■ Mathematics ■ Natural sciences[a] ■ Social sciences[a]		■ Government ■ Users ■ Public	
2001	4, 6, 4M	■ Language ■ Mathematics			
2004	2, 6, 3M, 5M	■ Language ■ Mathematics ■ Civics			
International assessments					
■ LLECE (1997) ■ PISA Plus (2001)					

[a]Administered to grades 4 and 6 only.

Uruguay

Both census- and sample-based tests have been used. They include multiple-choice and open-ended questions. The tests are complemented by questionnaires on the socio-educational context to be completed by teachers, principals, students, and families; the information thereby collected is used to analyze performance-related factors. The questionnaires focus on both in-school variables (such as infrastructure and facilities, human resources, teaching experience, management, and pedagogical concepts) and out-of-school factors (such as housing conditions, family composition, material and cultural goods, and parents' levels of education and occupation).

Table A-5: Uruguay—Summary of national educational assessment system

Test year	Grades tested	Subjects tested	Test coverage	Results reported to	Results used for
Domestic assessments, administered by Unidad de Medición de la Calidad de la Educación					
1996	6	■ Language ■ Mathematics ■ Student attitudes	National census	■ Government ■ Users ■ Public	■ Teacher training ■ Targeting support to students and schools
1998	3	■ Language ■ Mathematics ■ Natural sciences ■ Social sciences	■ National sample ■ Self-administered in all schools	■ Users ■ Public	Teacher training
1999	6	■ Language ■ Mathematics ■ Student attitudes	■ National sample ■ Experimental sample ■ Self-administered in all schools	■ Government ■ Users ■ Public	
1999	3M	■ Language ■ Mathematics ■ Natural sciences ■ Social sciences	National census		Targeting support to students and schools
2001	4	—	—	—	—
2001	Pre-school, 1, 2	■ Cognitive and affective development ■ Language ■ Mathematics	National sample	■ Government ■ Users	Teacher training
2002	6	■ Language ■ Mathematics	■ National census ■ National sample	—	—
International assessments					
PISA (2003)					

PREAL

Partnership for Educational Revitalization in the Americas
Programa de Promoción de la Reforma Educativa en América Latina y el Caribe

PREAL was established by the Inter-American Dialogue in Washington, D.C. and the Corporation for Development Research (CINDE) in Santiago, Chile, in 1995 as a multiyear initiative to build a broad and active constituency for education reform in many countries. It has become the leading non-governmental voice on education in Latin America and a strong advocate for involving leaders from civil society in education reform. Most of PREAL's activities are implemented by a region-wide network of expert public policy and research centers working to promote education reform.

PREAL seeks to improve the quality and equity of education by helping public and private sector organizations throughout the hemisphere promote informed debate on education policy, identify and disseminate best practices, and monitor progress toward improvement. PREAL's activities are made possible by the generous support of the American people through the United States Agency for International Development (USAID), by the Inter-American Development Bank (IDB), the GE Foundation, the International Association for the Evaluation of Educational Achievement (IEA), and the World Bank, among others. The contents of this publication are the responsibility of the author and do not necessarily reflect the views of PREAL or any of its donors.

INTER-AMERICAN DIALOGUE

Inter-American Dialogue
1211 Connecticut Ave., NW, Suite 510
Washington, D.C. 20036 USA
Tel: (202) 822-9002
Fax: (202) 822-9553
E-mail: iad@thedialogue.org
Internet: www.thedialogue.org & www.preal.org

CINDE

Corporación de Investigaciones para el Desarrollo
Santa Magdalena 75, Piso 10, Oficina 1002
Santiago, Chile
Tel: (56-2) 334-4302
Fax: (56-2) 334-4303
E-mail: infopreal@preal.org
Internet: www.preal.org