

## No. 11

### Educational Assessments in Latin America: Current Progress and Future Challenges

by Laurence Wolff\*

June 1998

*Laurence Wolff is an education consultant at the Inter-American Development Bank in Washington, D.C. Previously he worked for 22 years at the World Bank, where he was involved in education projects in Latin America, Asia and the Middle East. He holds a doctorate in education from Harvard, and wrote this report while a visiting scholar at the Education Development Center, a non-profit education research and development institute headquartered in Newton, Massachusetts.*

---

## Table of Contents

### I. Introduction

### II. Recent Advances in National and International Educational Assessments

International Programs	The French Experience
Experiences in the United States	Developing Countries

### III. The Latin American Experience

Chile	Mexico
Costa Rica	Argentina
Colombia	UNESCO/Orealc
Brazil	Regional Assessment Program

### IV. Lessons Learned and Future Challenges

Summary of Experience in Six  
Countries  
Consensus Building and  
Commitment  
New Approaches

Uses of Assessments  
Capacity Building and Technical  
Competence

Bibliography

Appendix: Assessment Systems in Six Countries at a Glance

---

## I. Introduction

Education is fundamental for economic, social, and cultural development, not to mention political stability, national identity, and social cohesion. Moreover, the high-tech businesses of today cannot thrive without people who can bring analytical, creative, and cooperation skills to the workplace. The availability of such workers can also have great impact on a country's ability to attract foreign investment. The economic accomplishments of East Asia, to take but one example, can be attributed in large part to the superior quality and level of education throughout that region, which now boasts of having four of the world's five best records in grade eight mathematics.

In recent years, nations throughout the world have come to agree on the importance of measuring educational performance. By assessing current levels of achievement and identifying obstacles to progress, they believe they can improve the type, depth, and breadth of education they offer. Educational assessments can be defined as measures of the degree to which curriculum goals, whether set by government authorities or national and international experts, have been achieved. National assessments evaluate the progress of institutions throughout the country. They differ significantly from the completion or entrance examinations designed to select students for another level of education. International assessments compare learning achievements across countries. Of course, measuring student learning will not by itself yield increased student achievement any more than weighing grain will yield increased agricultural output. It is, however, a necessary condition to establishing quantitative targets, assessing the tradeoffs of alternative resource allocation strategies and input combinations, and allocating resources and effort to achieve established targets. To ensure that educational assessments not only improve student learning but also are cost effective requires commitment, adequate financing, technical knowledge, managerial know-how, and political savvy.

Educational assessments are only one means of monitoring progress toward achieving educational goals. It is also important to assess quantitative outputs (e.g., the number of students enrolled, completing a given level, or being promoted); the adequacy of inputs,

such as textbooks, teachers, teacher-student ratios, and teacher training; classroom interactions and pedagogy; and performance in the labor market (e.g., how many graduates get jobs, and at what salaries).

Latin America is no stranger to educational assessments. In the late 1970s the *Programa de Estudios Conjuntos para la Integración Económica Latinoamericana* (Program of Joint Studies for Latin American Integration, ECIEL) completed a comparative study of learning in five countries using instruments developed by the International Association for the Evaluation of Student Achievement (IEA). Several Latin American countries have participated in international assessments sponsored by the IEA as well as by the Educational Testing Service (ETS). The results of these assessments have been disappointing, however; the countries of Latin America have consistently scored well below those of North America, Europe and Asia. In the IEA study of 1989 (Table 1), Venezuela scored below all other participating countries, including Indonesia. Trinidad and Tobago did somewhat better than Venezuela, but was still far below the developed countries. In a 1991 study of mathematics (Table 2) implemented by ETS, the cities of Fortaleza and Sao Paulo, Brazil, scored below all other participating countries and cities except Mozambique. As noted later in this report, Colombia, the only participating Latin American country, has also scored poorly in the most recent IEA mathematics study.

By 1991, Costa Rica, Mexico, Chile and Colombia had assessment systems in place. Since then, nearly every Latin American country has initiated a program of some kind. Under a grant from the Inter-American Development Bank (IDB), UNESCO has supported a regional program of testing third and fourth graders in reading and mathematics. In addition, international assessments have increased in depth and in the number of countries participating. This report examines advances in educational assessments both at the international level and in six countries of Latin America; it also comments on some possible new directions for assessment policies and programs in the region. The discussion draws on recent international and national studies (identified in the bibliography), including several prepared for a PREAL conference in Rio de Janeiro in December 1996.

## **II. Recent Advances in International and National Educational Assessments**

### ***International Programs***

The most important event in international assessments in recent years has been the IEA's Third International Mathematics and Science Survey (TIMSS). The IEA is a worldwide assessment consortium with headquarters in Amsterdam, in the Netherlands. It is known for its international studies of mathematics, science, reading, literacy, and social studies programs. IEA programs are financed by the participating countries.

The TIMSS seeks to measure, compare, and explain learning in science and mathematics in 41 countries. Mathematics and science examination results for children in grades four, seven and eight, and twelve are now available. In addition, the TIMSS recently completed an innovative analysis of children's "opportunity to learn," which categorizes and compares the curriculum, textbooks, and classroom pedagogy among the 41 participating countries. Such analyses demonstrate that achievement examinations need not simply concentrate on identifying countries with high or low scores but can be used as a tool to measure a country's educational progress, re-define its curricular goals, and change classroom practices.

In 1991, eleven Latin American countries (Costa Rica, Peru, Argentina, Dominican Republic, Colombia, Guatemala, Venezuela, Chile, Ecuador, Brazil and Mexico) attended a preliminary TIMSS regional meeting, but only Colombia and Mexico participated to the end. Furthermore, only Colombia's scores were reported in the TIMSS publication. At the last minute Mexico decided not to permit reporting of its scores. Argentina and the Dominican Republic participated in the curriculum analysis. The reasons for the low participation of Latin American countries (and of all developing countries) appear to be a combination of inadequate technical and financial resources and perhaps a misplaced desire of some authorities not to end up with the lowest scores.

The TIMSS scores for mathematics achievement in eighth grade are presented in Table 3. Among the 41 countries that reported their scores, Colombia ranked second last, just ahead of South Africa. The data show that only 4 percent of all Colombian students scored in the top 50 percent of students in the world; and, strikingly, none of the students in the Colombia sample scored in the top 10 percent in the world. The vast majority of countries that participated were developed countries with much higher per capita incomes than Colombia, and most spent significantly more on their education system. At the same time, countries with similar per capita incomes—notably Bulgaria, Lithuania, the Philippines, Romania, Latvia, Iran, Slovakia, Russia and Thailand—still scored higher than Colombia.

The TIMSS methodology enables curriculum developers and policymakers to determine whether curriculum, textbooks, and classroom teaching and learning are in alignment. As a result, the details of how Colombia fared on a variety of mathematics subjects, as shown in Table 4, provide a good idea of the strengths and weaknesses of the mathematics curriculum. Colombians are therefore able to use their performance to compare the quality of their educational objectives, curricula, and textbooks with those of other nations and thereby identify areas requiring improvement.

Although the main TIMSS program is nearing completion, a number of follow-up activities are already under way. TIMSS researchers have begun to create "benchmarks" for existing country curricula and tests on the basis of the TIMSS examination and analyses. In addition, there is now agreement that TIMSS will be replicated in 1999.

### *Experience in the United States*

The United States has taken the lead in developing a wide variety of educational assessments to evaluate institutional performance, establish minimum competency levels for school completion, and set benchmarks for measuring educational performance against international standards. The extent, complexity, and perhaps excesses of U.S. assessment programs can be seen in Box 1, which provides information on four assessment programs in Montgomery County, Maryland. The first is organized around a test developed by the county to provide parents with information on how well their children in grades three to eight are mastering the key elements of the curriculum at their grade level. It serves as a diagnostic tool to identify a student's strengths and weaknesses and uses multiple-choice, open-ended, and performance-based questions. The second program, the Maryland School Performance Assessment program, measures how well individual schools at grades three, five, and eight are meeting standards of performance set by the state. Its questions are mainly of the multiple-choice variety or require only short answers. The third assessment program, the Maryland Functional Tests, is a "minimum competency examination" in reading, mathematics, writing, and citizenship that all students must pass to receive a valid high school diploma. The fourth is the newly launched Comprehensive Tests of Basic Skills program, to be administered to a small sample of students in grades two, four and six. It will serve as an indicator of levels of basic skills in relation to national and international norms.

By history and law, and in contrast to most countries, the United States has no nationally mandated curriculum. And since most states provide only broad guidelines, individual school districts have considerable leeway in setting the curriculum. According to a recent report by the U.S. National Research Center (1997), state curricula are often poorly defined and disregarded at the local level. As a result, what is actually taught in the classroom varies greatly from school to school as well as district to district, and students and teachers often have an exaggerated notion of how well they are performing. Authorities have therefore mounted an effort to establish stronger state standards for learning and achievement, as well as possibly voluntary national standards. Similarly, a number of states—Maryland is one—are seeking to base their curricula and tests on international standards, especially as related to TIMSS. It should be noted that the two major teachers unions, especially the American Federation of Teachers (AFT), have consistently supported state and national efforts to improve educational assessment and quality.

### ***The French Experience***

Since the mid-1980s France has developed a sophisticated student assessment system. While France has a nationally mandated curriculum, elementary and secondary schools have a great deal of freedom in the use of their funds and their pedagogical techniques. Responsibility for monitoring and assessing education rests with the more than 200 technicians in the *Direction de l'Evaluation et de la Prospective* (DEP), which is a division of France's Ministry of Education. The highest government authorities are strongly committed to building a system of measuring educational outcomes that is based

on clear standards and full and open dissemination of results, thus ensuring the independence of the DEP.

In the French system, achievement standards are centrally set and tests are given annually to all students in specific primary and lower-secondary grades. The baccalaureate (completion) examination administered to upper-secondary students is also used for assessment purposes. The DEP provides detailed, user-friendly feedback to all parties concerned: the general public, parents, students, teachers, and administrators, as well as the minister of education. The agency has also developed a wide variety of optional teaching and learning materials tailored to the performance of students and the strengths and weaknesses of teachers. As various agencies have done in the United States, the DEP has sought to rate school performance by international standards.

As is the case throughout the world, the schools in France's poorer communities score lower than those in richer ones. The DEP has sought to identify schools in poverty-stricken areas that are performing well in order to define better the "value added" of schooling. One statistical approach to rating the performance of schools in relation to the socioeconomic status of their students is shown in Graph 1. Schools doing better than might be expected of their student population are termed "effective" schools. By studying such schools, authorities should be able to determine why they are effective and then establish similar conditions in poorer-performing schools. This kind of analysis is already being done in some parts of the United States (see, for example, a description of "value added" in the Dallas school system as described in Alvarez and Ruiz-Casares, 1997).

### *Developing Countries*

A move to establish national systems for assessing education is also under way in developing countries. Outside Latin America, assessment systems have been described and analyzed in Thailand, Egypt, South Korea, and Jordan, among others. Though small and relatively poor, Jordan has developed a systematic, technically competent assessment program. Its success in this regard can be traced to: (a) the direct involvement of the highest authorities (e.g., the crown prince) and their long-term commitment to improving the quality of education and to open reporting of results; (b) the establishment of an independent, well-financed agency outside the Ministry of Education responsible for assessment; and (c) strong technical leadership, coupled with assistance from abroad. In 1991 Jordan carried out several rounds of testing that were based on international assessments developed by the Educational Testing Service's IAEP test, and it is now providing technical assistance in this area to neighboring countries. A recently completed round of assessment has identified significant improvements in student learning as a result of Jordan's decade-long efforts to reform school curricula and teacher training.

## **III. The Latin American Experience**

Before 1991, Chile and Mexico were the only countries in the region with much assessment experience. Chile's assessment program had been in operation since 1980, and although Mexico's had also been in existence for some years, the authorities made little or no effort to disseminate the results. Costa Rica undertook national assessments between 1986 and 1990. Colombia had a long-standing national testing system mainly used for selection into higher education.

In the next few years, other Latin American countries also began moving in this direction. By 1996 their experience had broadened considerably and almost every country of Latin America has now initiated an assessment program of some kind. This report draws conclusions from developments in Chile, Costa Rica, Mexico, Argentina, Brazil and Colombia. It also describes the UNESCO/Orealc regional program.

### *Chile*

As already mentioned, Chile has long been involved in educational assessment. Its program was conceived in 1978, when the Ministry of Education asked the Pontificia Universidad Católica to design and implement an information system for education. In 1988, with the transfer of public schools to municipalities, the program was renamed the National Program to Measure the Quality of Chilean Basic Education (SIMCE).

The function of SIMCE is to help the Ministry of Education and regional and provincial authorities supervise the education system, evaluate individual schools, and assist in teacher in-service training. The program tests children in grades four and eight in Spanish and arithmetic and 10 percent of them in the natural sciences, history and geography. It also assesses personal development and attitudes, the attitudes and background of teachers and parents, and school efficiency. Assessments of the two grades take place in alternate years. Beginning in 1991 the Ministry of Education took full charge of administering the program.

Since 1988 the program has become more effective and efficient, following improvements in technical capacity, computerization and administration. Scores are now delivered to schools more rapidly, and reports have been simplified to ensure that the results are easily understood. These efforts appear to be having an impact on pedagogical planning in many schools. Those in charge of designing curricula and instructional materials, for example, are emphasizing that objectives be mastered whenever students appear to be having problems. Some parents are using SIMCE results to select better-performing schools for their children. The cost of the program is about US\$5 per student, which is comparable to international standards.

SIMCE scores have also formed the basis of a pedagogical program directed initially at 900, and later, 1,200 of Chile's poorest-performing schools. The schools under this program have been provided with educational materials, libraries, books, infrastructure, and in-service training. Depending on their initial condition, certain schools are given preference in the award of grants for local improvement programs. In addition, schools increasing their scores from year to year have received financial rewards.

SIMCE tests reveal the following: schools with children from poor and uneducated families or from rural areas have the poorest scores on SIMCE tests; public municipal schools and rural schools score worse than private schools, especially the long-established institutions; private schools perform somewhat better than public schools even after controlling for the socioeconomic status of parents; and finally, scores in the 900 schools appear to have improved significantly in recent years. Current problems include the fact that, while many schools are making use of SIMCE results to improve local conditions, there is still an expectation that remedial action should be initiated by central authorities. Some schools have reported a rise in the number of students from deprived circumstances in an apparent effort to show that their relative (value-added) achievement has improved. Measurements of the affective domain have not been successful and should perhaps be abandoned. It may also be appropriate to reduce the amount of universe testing and to rely more on sampling. Finally, because of technical problems, comparability of results from year to year is inadequate.

Overall, Chile now has the most comprehensive and best-managed assessment system in Latin America, and SIMCE has served as a strong tool for implementing a reform program fostering decentralization, accountability, and increased learning. Government authorities are seeking to further improve the system. In particular, they are now planning to add performance testing to the assessments (currently all tests are of the multiple-choice variety). Since the causes of poor performance are still not fully understood, a more sophisticated research effort is underway. Finally, Chilean authorities have decided to participate in the follow-up TIMSS study scheduled for 1999 as a means of benchmarking the performance of their students with the rest of the world.

### *Costa Rica*

Costa Rica has been involved in educational assessments since 1986. Its activity in this area has progressed through three stages. During the first stage—from 1986 to 1990—the stated objectives were to measure the extent to which children and young people were learning basic concepts, to encourage parents and teachers to use teaching time more effectively, to stimulate a national discussion of the quality of education, to point out that all Costa Ricans are responsible for improving the quality of education, and to demonstrate the importance of reestablishing national certification exams at the end of secondary school.

The program was formulated in large part by the Institute for Research to Improve Costa Rican Education (IIMEC), an autonomous institution of the University of Costa Rica. Initially the program introduced tests in Spanish and mathematics for all students in grades three, six, nine, eleven and twelve. In 1987 and 1988 the tests were expanded to physical science, social science, English and French. In 1988, a secondary school completion examination was established. These tests led to public controversy because of the low level of student achievement, and comprehensive objective testing was scaled back.



From 1988 to 1993, the focus was on the secondary school completion examination. The tests, prepared by IIMEC and implemented by the Ministry of Education, counted for 60 percent of each student's final grade and were scored at the local level. When a 1991 study found that 30 percent of the tests were scored incorrectly, in almost all cases in favor of the student, the authorities decided to eliminate most of the performance questions and to score nearly all the tests centrally, using optical readers. In addition, sixth-grade exams, prepared by regional authorities, counted for 50 percent of those students' final grades, but little effort was made to ensure that the results reported by regional authorities were reliable or valid.

Beginning in 1993, under a new government and minister of education, Costa Rica firmly committed itself to assessment and received help in this regard through a project financed by the World Bank. Diagnostic tests were reestablished for grades three, six and nine in four subjects; other types of assessments were undertaken; and IIMEC was strengthened with new equipment and personnel. The new system administers a wide variety of tests and assessments, most of them prepared by IIMEC.

Among the formative assessments are:

- 1.) *Diagnostic achievement tests* in the basic subjects for grade three. These were pre-tested in 1995 and administered to a large sample in 1996.
- 2.) *Initial diagnostic assessments*, given to 10 percent of the children entering first grade to obtain information on their physical, cognitive, and social-emotional status. The results are to be used to establish guidelines for the provision of appropriate learning experiences in preschool and first grade.
- 3.) *Tests of problem-solving skills*, given in 1996 to a sample of ninth graders. They were designed to measure cognitive capacities in this area as well as socioeconomic and academic variables influencing performance.
- 4.) *Evaluations of physical capacity*, conducted in 1996 to measure the physical aptitude of a national sample of children in grades three, six, and nine.

Several summative assessments have also been undertaken. A completion examination prepared at the national level is given to all ninth graders in all the basic subjects. Multiple-choice questions are graded by optical readers, while performance questions are graded by specially trained teachers. The results count for 25 percent of the second semester grade. Since 1988 Costa Rica has also administered a *secondary school completion examination*, which is prepared and given regionally. It counts for 60 percent of the final qualification and contains both objective and performance questions. In addition, *sixth-grade completion examinations* are prepared and given by the regional Ministry of Education offices, but they vary in quality, validity and reliability.

In the initial testing (1988), children performed far below the expectations of the national curriculum. Urban private schools did best on these examinations. Although the IIMEC

distributed summary information to each school, the institutions made no explicit attempt to utilize the results. However, a 1989 survey of teachers showed that 70 percent were aware of the material, and about 35 percent used it in the classroom in one way or another.

In the second stage, information from the secondary school completion examinations was made available to individual students and to individual secondary schools. How the information was used was left entirely up to the school.

The third stage is expected to show a more systematic approach to dissemination, utilization, and feedback. Reports will be prepared for different grade levels, and regional education units will be asked to incorporate assessment results into their educational plans. An attempt will also be made to measure the extent to which education has improved. Despite these advances, parents receive no information other than student scores, there are no special press releases on education, the results are not used to prepare specific materials to correct weaknesses in learning, and training institutions have received no specific guidance or information on the tests. However, reports will be provided to individual schools, especially to those concerned with first-grade and preschool activities, accompanied by suggestions for enriching the curriculum.

On the whole, Costa Rica's assessment efforts have been hampered by a stop-and-go approach. Although the previous minister of education was highly supportive of the program, some educators and political leaders are not, and it is unclear whether this situation will continue under the new administration, which took office in May 1998. Given its small size, Costa Rica may well have embarked on too ambitious a program. A more productive course of action might be to reduce the number of tests and to concentrate on disseminating the results and putting them to use. Costa Rica's "high-stakes" secondary examination, which counts for 25 percent of the final grade as a means of raising achievement at this level, may be of interest to other Latin American countries.

### ***Colombia***

In 1986, the government of Colombia established a department in the Ministry of Education that was responsible for assessing and evaluating education institutions and programs. Then, in 1990, it initiated a national assessment system with a view to constructing an operational model for assessment that would lead to decisions that could improve quality. Furthermore, the system was expected to evaluate student knowledge and determine how teachers, schools, and educational materials promoted it; to generate a greater sense of the importance of schooling; to communicate this sense to all of society; to provide the theoretical background for concepts such as quality and assessment of educational achievement; and to support research that would help the country achieve these objectives. The program was undertaken by an inter-institutional team composed of staff of the Ministry of Education, the National Teachers University, the Center for Social Studies, and the Instituto SER.

Since 1968 the Colombian Institute for the Enhancement of Higher Education (ICFES) has been responsible for formulating a national examination for entrance to the nation's institutions of higher learning. In 1980 these examinations became obligatory and thereafter were given by the National Testing Service of ICFES. While they focus on student testing and entrance to institutions of higher education rather than on assessing the educational system as a whole or its subsystems, the program has helped to create technical competence in testing.

Unlike Chile and Costa Rica, Colombia emphasizes sampling and research as a means of identifying the causes of low achievement. Therefore it does not provide information to every parent, student, or school about their performance. Colombia is now considering expanding its sample to ensure that feedback is available to larger numbers of municipalities and schools.

The Colombian system began with a test in mathematics and Spanish for students in grades three and five. Tests in the natural and social sciences were subsequently developed for grades seven and nine. These tests explicitly measured higher-order reasoning skills, such as the use of algorithms and problem-solving skills in mathematics and the ability to extract meaning from written Spanish.

When the cognitive achievements of students in grades three and five are measured against the national curriculum, the results are disappointing. Few students are able to perform the basic operations required to solve concrete problems, and few fully comprehend what they read, either in a critical or reflective sense. The highest scores are achieved in urban and private schools and in certain regions. Even in a study controlling for socioeconomic status, attendance in private school appeared to confer an advantage in terms of achievement. Achievement was higher among those who attended preschool, did not repeat a grade, were seldom absent, had books in their homes, and had parents with higher levels of education. If textbooks were available in school and were used, achievement was higher still. Students with better-trained teachers, more textbooks, female teachers, and attending complete schools also did well. The tests also confirmed that students in the *Escuela Nueva* ("New School"), an innovative program for small rural schools, performed significantly better than rural school students who were not in the program.

The conclusions of these assessments have influenced the objectives and content of the new education law passed in 1994. However, the information gathered through the current sample-based assessments is of little use to specific municipalities and schools which have become increasingly responsible for their schools under the new decentralized education system. Although present authorities think it would cost too much to expand assessment to all students, Colombia plans to work with a larger sample and hopes to establish a data and question bank that can be used directly by schools and education authorities.

As mentioned earlier, Colombia is the only country in Latin America to have participated fully in the TIMSS program. Although the results were disappointing, Colombia should

be applauded for its willingness to accept outside scrutiny. Colombia is now seeking to utilize the TIMSS results to strengthen the national assessment system and to reform the curriculum.

Colombia's assessment program has had a strong research element which has affected overall policy, notably in the writing of the 1994 Education Law and in confirming the success of the *Escuela Nueva*. However, the results have not yet been used systematically to improve the performance of individual schools or to reform the curricula. The mathematics and science curriculum, in particular, are in need of reform. At the same time, the partnership between public and private institutions has worked well.

### ***Brazil***

Brazil was among the late starters in developing assessment tools at the national level. It established the National System of Evaluation of Basic Education (SAEB) in 1990 and only began taking samples in 1993 and 1995. One of SAEB's objectives has been to encourage states and municipalities to initiate their own assessments. The states of Paraná, Minas Gerais, and Brazil, in particular, have recently initiated assessment programs.

In 1995, SAEB announced a number of innovations. Its survey would include both secondary education and private institutions. It also adopted more sophisticated methods of measurement, introduced instruments that would provide information on student background, and reduced the turnaround time for the publication of results. The 1995 survey focused on grades four and eight in basic education, and on grade three at the secondary level.

Survey results, published in 1995, indicate that 90,499 students were tested in primary grades four and eight and secondary grades two and three. This sample of students was drawn from 2,289 public and 511 private schools. Survey items were based on the level of learning expected by teachers and education specialists. In reading, students were tested for understanding, extension and critical examination of meaning. In mathematics, the survey focused on three categories: comprehension of concepts, understanding and application of procedures, and problem solving. In the statistical analysis, the main task was to determine the expected and actual levels of performance for children at various grade levels.

In the 1995 test, children throughout the country scored significantly below the levels expected by teachers and specialists. In mathematics, only 21 percent of the students in grade four scored above the expected level, only 15 percent in grade eight scored at or above, and only 4 percent of secondary students scored above. Language scores were even lower. Only 22 percent of fourth graders, 14 percent of eighth graders, and 1 percent of secondary school students scored above the expected level. The highest scores were reported in the south, southeast, and center-west of the country. Scores were lowest in the north and northeast. Students in the major cities scored better than those in the interior.

Children with more educated parents and those attending private schools also scored higher. Students attending night schools scored lower. Older students performed worse than younger students. White and Asian students scored higher than those of mixed or black background. And students with more highly trained teachers generally scored higher than those with less well trained teachers. These results have important implications for educational policy, but authorities have only just begun discussing them. Furthermore, they have yet to incorporate more sophisticated techniques in their analyses, such as controlling for the socioeconomic status of entering students.

Recently the state of Sao Paulo began a comprehensive assessment program. All students in selected elementary and lower secondary grades are tested in mathematics and language, and the results are sent to parents, teachers and schools. An analysis of items where students have scored lowest is provided. "Anchor items" are included to ensure compatibility of results from year to year. The Sao Paulo program is undertaken mainly under contract with private, non-profit testing agencies. The Minas Gerais assessment program also appears to be making progress in the utilization of its results since individual scores are being reported to schools.

Quite separately from the above, under a law passed in 1995 the Ministry of Education has developed a system for assessing institutions of higher learning. The purpose is to inform students and society in general about the quality of higher education institutions. Schools with low scores will be required to devise programs to strengthen quality that would be supported by the federal government.

To summarize, the Brazilian assessment program has started only recently. Brazilian authorities are only now beginning to consider how assessments can be used to improve educational policy and the curriculum. There is a need to conduct more rigorous analyses to identify factors affecting achievement. Since education in Brazil is highly decentralized and follows no national curriculum, the government is considering establishing some voluntary national standards and a test to measure whether they have been met. The higher education assessment program, the first of its kind in the region, is an important innovation.

### *Mexico*

In 1970 Mexico established an office in the Education Planning Unit of the Secretariat of Public Education, which eventually became the Sub-Directorate of Evaluation and Accreditation, to examine the characteristics and quality of the country's education system. The staff subsequently tested the aptitude of children in grade six of basic education and established an examination for entrance to secondary schools.

From 1976 to 1982, the sub-directorate investigated learning in a representative sample of fourth and fifth graders. The results of this assessment appeared in scientific and scholarly publications but otherwise were not made public, and the authorities paid little attention to them. In fact, assessment information became a "state secret" known only to a

small number of secretariat staff. This approach hampered technical development and policy utilization.

During the period 1983-1988, Mexico developed an examination for graduates of teacher training schools. Then, in 1989, it decided to apply the concept of assessment more widely to improve teaching and learning, and to publish the results. In 1992, the federal government and the National Teachers Unions agreed on a program to modernize basic education by decentralizing it to the state level, but leaving the federal government to measure and evaluate learning and to ensure the quality of basic education and teacher training. To this end, the Secretariat of Public Education committed itself to supporting teacher, classroom, and national assessments.

In 1994, after five years of assessing the quality of education in Mexico, the secretariat released a report on the knowledge and skills of 480,000 teachers and achievement of 2.8 million children at the primary and secondary levels. Its principal conclusions were that children who attended preschool scored higher than those who did not; children repeating sixth grade or working did worse than their counterparts; children attending urban or private schools did much better than those in rural and public schools; those who achieved the lowest scores were in indigenous and community schools with poor facilities and less highly trained teachers; and those who scored highest were in urban schools and had more highly educated parents. Although children in grades one and two scored close to what researchers and curriculum developers expected, their scores as a percentage of correct answers went down in successive years. Mexican authorities also reported that it was impossible to measure systematically the classroom performance of teachers because student populations are extremely diverse and technical difficulties still abound.

As mentioned earlier, Mexico participated fully in the TIMSS but at the last moment decided not to release the results. This decision should not be so surprising in view of Mexico's general reluctance in the past to disclose exam results, which at times were treated as state secrets. Even though attitudes have changed since then and the results of elementary and secondary assessments are now made public, reporting is still done with some ambivalence. The most distinctive feature of Mexico's assessment system is that it has systematically tested teacher knowledge and capacity.

### *Argentina*

Like Brazil, Argentina is a late starter in assessment, but it has moved more rapidly to establish and utilize assessments for improving educational quality. Under a new federal law decentralizing education, the Ministry of Education in 1993 established a national evaluation system, with offices in the Secretariat of Programming and Educational Evaluation. The objectives of the evaluation system are to promote decentralization, provide key information on the status of education, monitor progress in achieving reform objectives, identify inequities and inadequacies and areas in which compensatory programs are needed for disadvantaged populations, and encourage broader sectors to

participate in education decisionmaking. Since 1993 Argentina has tested all children in the last year of primary and secondary school in language and mathematics skills. In 1995 the assessments were expanded to cover grade three at the primary level and grade two at the secondary level. Tests in social and natural sciences were also initiated. Complementary questionnaires were given to teachers, directors, students and parents. The tests were based on a detailed analysis of expected curricular achievement and careful pilot testing. The reliability and validity of these tests were confirmed through statistical analysis and expert review of the extent to which items correspond to curriculum objectives. The entire program is managed by the Ministry of Education. Assessment results are incorporated into a larger system of education information that is used to monitor and supervise education at all levels.

Assessment information includes details on school performance and profiles of students, teachers, and management models associated with performance. This information is disseminated to national, state, local, and school authorities. However, individual school scores are not reported to schools. Several reports have also dealt with the subject matter content and include recommendations for improvements in pedagogy, both centrally and at the school level. Test results have also been used to develop manuals on how to improve learning as well as in-service training, and these guides, along with technical assistance, have been provided to schools in all 24 of the country.

Although test questions were geared to minimum expected levels of response, on average students answered only 50 percent of the questions correctly. School factors account for 40 percent of the variance associated with mathematics scores and for 28 percent of the variance in language scores. Children from families at the top of the socioeconomic ladder scored highest. However, differences between schools were greater than differences within schools, which suggests that Argentina's education system is highly segmented. Poor children in heterogeneous classes scored higher than poor children in homogeneous classes. Overall, the results show that there is room for significant improvement within schools.

Though initiated only in 1993, Argentina's assessment program is well thought out and strongly linked to its strategy for decentralizing and improving the quality of education in general. Like Chile, Argentina is fully intent on making use of its assessment results. It has already given wide distribution to national and regional reports, many of which address significant problems as well as provide pedagogical materials designed to strengthen teaching.

### ***UNESCO/Orealc Regional Assessment Program***

Through a grant from the IDB, the regional office of UNESCO (*Oficina Regional para América Latina y el Caribe, Orealc*) is implementing a regional assessment program called the "*laboratorio latino-americano de medición de la calidad de la educación*" (Latin American educational quality measurement laboratory). The program, initiated in 1995, seeks to encourage coordination and strengthen the capacity of national assessment

agencies in the region. In 1997, mathematics and reading tests, accompanied by detailed background questionnaires, were given to a sample of third and fourth graders in 15 Latin American countries. The tests were developed by a committee of all participating countries and are based on informal review of curriculum objectives in mathematics and reading in the region. The results will be reported around mid-1998.

In 1993, UNESCO implemented a similar test, on a pilot basis, given to fourth grade students in seven Latin American countries. Most students scored far below the minimum levels expected by regional curriculum and testing experts. On average students were able to answer around half of the questions correctly. Venezuela and Costa Rica scored highest while the Dominican Republic, Bolivia and Ecuador scored lowest. Table 5 summarizes the results.

UNESCO is planning to commission a study comparing the degree of difficulty and the curriculum content in the UNESCO mathematics test with those of the TIMSS mathematics test. This will provide indirect evidence of how well Latin American countries might do compared with countries outside the region. TIMSS experts could also assist selected Latin American countries in implementing the TIMSS methodology for measuring "opportunity to learn"—e.g., reviewing and comparing official curriculum, implemented curriculum, textbooks, and classroom pedagogy.

The UNESCO program is an important step in establishing regional coordination. To ensure the long-term strengthening of regional capabilities, future efforts require supporting regional centers of excellence, linking public and private institutions, and maintaining stronger links with IEA and leading world centers of research and development than in the past.

#### **IV. Lessons Learned and Future Challenges**

##### ***Summary of Experience in Six Countries***

Chile and Argentina have the most comprehensive and best-managed assessment systems in Latin America. Chile has also demonstrated the strongest long-term commitment to assessments. Although Chile and Argentina have made the greatest strides in using assessments for policy purposes, curriculum reform, and improvement in individual schools, there is still much room for improvement even in these countries. In particular, clear national learning objectives have not yet been set and systematic efforts at "aligning" curriculum, textbooks and classroom pedagogy have just begun. Chile and Argentina have also made some progress in integrating assessments into large, monitoring and evaluation systems.

Among the key problems that require greater attention are Mexico's reluctance to disseminate assessment results; Costa Rica's stop-and-go approach to assessment, and its excessive number of tests; and Brazil's slow progress in utilizing assessments for curriculum and policy reform. With the exception of Costa Rica, the six countries



discussed here have relied on multiple-choice rather than performance-based or open-ended testing.

Another point to note is that all six countries reported similar assessment results. In particular, students in the later years of primary school and those in secondary school scored far below the expectations of professional educators and researchers. Students from urban and private schools and with more educated parents had the highest scores. Few countries in the region have yet attempted to conduct detailed multivariate analyses of these results to determine their causes. Although several studies have shown a correlation between better-trained teachers and higher student performance, most did not adequately control for the socioeconomic status of students or school location to confirm the importance of teacher education. In addition, little is yet known about whether student achievement levels have changed over time.

Notable innovations in the region include Brazil's attempt at assessing institutions of higher learning; Costa Rica's assessments of learning readiness among children entering primary school and its examination for students completing secondary school; Mexico's tests of teacher knowledge and skills; Colombia's emphasis on research; Chile's use of assessments for targeting resources; and Argentina's efforts to use assessments for curriculum reform. The UNESCO/Orealc program is also an important innovation which will have a greater impact as coordination with institutions like the IEA is increased.

Overall the key lessons learned and challenges for the future include: (a) the importance of national consensus and long-term commitment; (b) the importance of explicitly focusing on the use of assessments as tools to improve learning; (c) the need for capacity building and technical competence; and (d) the need to benefit from developments and innovations coming from both within and outside Latin America, especially through international testing programs like those of the IEA.

### ***Consensus Building and Commitment***

#### ***• National Consensus Building***

Perhaps the most important lesson to date is that the countries of Latin America no longer need to debate whether or not to perform assessments. Instead they need to set educational goals, determine whether children, institutions, and school systems are meeting those goals, and then establish programs to ensure that these goals are eventually met. Assessments will not serve to improve the quality of education unless everyone agrees on the importance of improving quality and on the full, timely, and user-friendly reporting of assessment results to all key stakeholders (i.e., "transparency"). This consensus must represent a coalition of teachers, parents, administrators, and business and political leaders. It must persist over the long term and have the backing of the highest government authorities, yet its decisions must not be influenced by political partisanship. Assessment programs must forge ahead and not be allowed to stop and start intermittently (as in Costa Rica) or keep their results from the public (as in Mexico).

### • *Integrating Teachers into the Assessment Process*

If assessments are to have an impact on education, teachers need to be included in the process from the beginning. Teachers unions throughout Latin America tend to be passive bystanders or even opponents of assessment. They must be persuaded that it is in their best interest to adopt a more professional approach to their work, upgrade classroom procedures, and participate fully in all assessment programs. Unions need not fear that the results will be used to reward or punish individual teachers, since the student body varies so much from one year to the next that the quality of teaching will always be difficult to measure accurately, as experience in France, the United States, and Mexico shows. If assessments are viewed as a means of evaluating schools and local school systems rather than teachers, then parents, teachers, and administrators will be better able to work together as a team to improve a school's achievement scores.

### • *A Systems Approach*

A commitment to transparency also requires a systems approach to monitoring and assessment. In this respect, it must be understood that assessments measure only the performance of children in school and tell little about the extent and causes of dropout. Other tools for measuring student and system performance include: studies of drop-out and repetition; measurements of school inputs and estimates of the minimum resources every school should have; systematic observations of school processes; and studies of labor market performance of school leavers.

### *Uses of Assessments*

The most important challenge for the future in Latin America will be to ensure that assessments are in fact used as tools for improving the quality of education. Assessments can influence many elements of the education system, including national education policy; education reform programs; education curriculum; decisions by parents, students, and teachers; local and regional education policy; and school-level pedagogy and programs. Three areas of particular importance in Latin America where assessments can be used are strengthening "alignment," targeting resources to the neediest schools, and encouraging responsibility and rewarding achievement at the school level.

### • *Strengthening "Alignment"*

Above all, assessments can help the countries of Latin America bring into closer alignment the intended curriculum, the actual curriculum, textbooks, teacher knowledge, classroom pedagogy, and learning. The resulting adjustments will make it possible to set higher but attainable goals, improve teacher training and pedagogy, and ensure that the revised curriculum is followed in practice. This will require a strong effort at analysis,

dissemination, and training. With ample feedback and resources, assessments can then be directly used as a tool to improve teaching and learning. To date Argentina has made the most progress in this process.

- ***Targeting Resources***

In view of the great disparity in educational attainment across Latin America, assessments should be used to channel additional resources to the region's neediest schools, particularly those in rural areas and in the urban slums. Such activity might be patterned along the lines of Chile's "900 schools program," which is explicitly directed at improving the worst-performing schools.

- ***Encouraging Higher Performance in Schools***

Assessments must provide useful information to schools, parents and teachers. This will require, among other elements, determining the "value added" of schooling, by identifying schools that score higher than expected given the socioeconomic background of their students, as is done in France, or by measuring the improvements in a school's scores over time. These approaches lead to recognition of and rewards for "effective" schools, and put pressure on other schools to replicate the same characteristics. This approach also prevents bias against schools in poor neighborhoods. Up to now Chile has done the most to recognize and encourage such schools.

### ***Capacity Building and Technical Competence***

- ***Strengthening Independent Institutions***

In most countries of Latin America, the central government would have great difficulty directly implementing an assessment program over the long term, as is done in France. Among the obstacles to such a system are potential conflicts of interest, inadequate salaries, and inflexible hiring and management procedures. A more appropriate strategy for the region would be to foster competent testing agencies outside the government but to maintain a small, knowledgeable, and highly efficient government agency to oversee their activities. Colombia and Brazil seem to have developed public-private partnerships of this sort.

- ***Ensuring Adequate Technical Expertise and Supporting Research***

Assessments require a high level of expertise if they are to succeed in measuring what they are intended to measure. Otherwise it is all too easy to fall short: they may not adequately discriminate among learners; comparability from year to year may be compromised; schools and regions might cheat; the sample framework may be inexact and render the entire assessment useless; and more sophisticated analytical techniques

may show that supposed causal relationships are spurious. Many of these problems have already been detected in country assessments. It is therefore essential to train and adequately remunerate experts in curriculum testing, sample-survey methodology, and analytical techniques. As the Jordanians would agree, it is also important to consult the world's best-known and respected assessment authorities, since there is no such thing as a "regional" testing methodology.

Governments and foundations also need to strengthen graduate centers of excellence in teaching and research in the region. In particular there is a need to support independent research based on data from assessments, which are a rich source of information on all aspects of education as it relates to students, teachers and schools. For the best results, this research should be based on up-to-date analytical techniques. Equally important, researchers and policymakers must agree on the goals of research and be willing to disclose the results, even if ambiguous or negative.

### *New Approaches*

#### **• *Participating in International Assessment Programs***

TIMSS offers the countries of Latin America an enormous opportunity to set standards, align curriculum and textbooks, upgrade teachers, and achieve breakthroughs in learning. The entire region may well benefit from TIMSS expertise and methodology.

The IEA is currently planning replication of TIMSS in 1999. Chile and Brazil are planning to participate. Latin American countries could also participate in the new IEA civic education program, which will be particularly important for issues related to strengthening civil society in the region. Currently Chile and Colombia are committed to the civic education study.

Alternatively, TIMSS researchers could review curricula in texts and curriculum guides, and compare them with curricula in other areas of the world. TIMSS researchers could also try to determine whether national and other examinations are in line with the intended national curriculum. TIMSS methodology has already been used to examine curriculum content, textbooks, and pedagogy in Argentina, Colombia, Mexico, and the Dominican Republic. TIMSS international staff and regional researchers could be asked to provide technical assistance in any of these activities.

#### **• *Universal versus Sample Testing***

Before countries undertake universal testing, they should be aware that this is a costly endeavor and requires close and capable management. A system of detailed feedback and assistance to individual schools is also needed. In Chile universal testing costs about \$5 per student, or about 2 percent of total unit costs. Even this small amount may be difficult to justify in countries facing severe fiscal constraints and lacking funds for educational materials.

By comparison, sample surveys cost less but can still provide adequate information for setting national policies and for identifying problems of alignment between the curriculum, textbooks, and teaching. However, these surveys require careful sampling and rigorous management. Wherever possible, the countries of Latin America also should use existing selection exams for assessment purposes.

- ***Performance Testing***

In recent years the United States has been incorporating more and more open-ended measures in its performance testing, while Europe has been concentrating on improving the reliability and validity of measurement instruments. In contrast, most of Latin America still favors the multiple-choice type of assessment. The countries in the region should at least explore the newer approaches to testing. They will be unable to do so, however, unless those who grade the tests are fully trained in the procedures.

- ***Minimum Competency Examinations***

The countries of Latin America may also wish to consider minimum competency exams for the secondary level. These have already been developed in parts of the United States and in Costa Rica as a means of encouraging learning. Examinations of this type would be especially appropriate for larger, decentralized countries, such as Brazil.

- ***Options for Smaller, Poorer Countries***

The smaller and poorer countries simply do not have the technical expertise or resources to mount a sustained assessment. As the experience of Costa Rica suggests, local expertise can all too easily be spread thin. Smaller countries should therefore aim for only a small number of tests (e.g., only in math and language) in a few grades. Some other options are to operate in consortia (composed, say, of Central American countries), ride piggyback on international programs, and take advantage of regional technical assistance (e.g., from Chile and Argentina).

- ***Regional Cooperation***

To strengthen the process of regional coordination, it will be necessary to develop closer links with the IEA and with leading world centers of research and development in testing and assessment, and to establish broad participation, including non-governmental institutions, in regional fora and committees. Regional efforts are also needed to: analyze the results of the UNESCO/Orealc assessments so as to identify factors associated with high achievement; understand the interactions between curriculum goals and learning through measuring "opportunity to learn"; and, most importantly, identify and support regional centers of excellence in testing, measurement, and educational research.

---

## Bibliography

- Alvarez, Benjamin, and Monica Ruiz-Casares, eds. 1997. *Assessment and Education Reform: Policy Options*. 1997. Washington D.C.: Academy for Educational Development.
- Capper, Joanne. 1996. "Testing to Learn, Learning to Test." Newark, Del.: International Reading Association.
- Esquivel, Juan Manuel. 1996. "Medición de logros del aprendizaje y el empleo de los resultados en Costa Rica." PREAL Seminar on Evaluation and Standards in Latin American Education, Rio de Janeiro.
- Horn, Robin, Laurence Wolff, and Eduardo Velez. 1991. *Developing Educational Assessment Systems in Latin America: A Review of Issues and Recent Experience*. Washington, D.C.: World Bank.
- Ministry of Education and Culture, Brazil. 1996. "Seminario Internacional de Avaliação da Educação," October 1995, Rio de Janeiro.
- de Moura Castro, Claudio, et al. 1984. "Determinantes de la Educación en América Latina: Acceso, Desempeño, y Equidad." ECIEL (Programa de Estudios Conjuntos para la Integración Económica Latinoamericana). Rio de Janeiro: Fundação Getulio Vargas.
- Murphy, Paul, Vincent Greaney, Marlaine E. Lockheed, and Carlos Rojas, eds. 1996. *National Assessments—Testing the System*. Washington, D.C.: World Bank, Economic Development Institute.
- National Research Council. 1995. *National Science Education Standards*. Washington, D.C.: National Academy Press.
- Rodríguez Schuller, Cristina. 1996. "Sistema de medición de la calidad de la educación: características y uso de los resultados de la evaluación para mejorar la calidad [Chile]." PREAL Seminar on Evaluation and Standards in Latin American Education, Rio de Janeiro.
- Schmelkes, Silvia. 1996. "Propuesta para el diseño y desarrollo del sistema nacional de evaluación integral de la educación básica en México." PREAL Seminar on Evaluation and Standards in Latin American Education, Rio de Janeiro.
- Schmidt, William A., Curtis McKnight, and Senta A. Raizen. 1997. *A Splintered Vision: An Investigation of U.S. Science and Mathematics Education*. Dordrecht: Kluwer Academic.

Schmidt, William A., et al. 1997. *Many Visions, Many Aims: A Cross-National Investigation of Curricular Intentions in School Mathematics*. Dordrecht: Kluwer Academic.

Schmidt, William A., et al. 1997. *Many Visions, Many Aims: A Cross-National Investigation of Curricular Intentions in School Science*. Dordrecht: Kluwer Academic.

U.S. National Research Center for the Third International Mathematics and Science Study. 1997. *A Splintered Vision: An Investigation of U.S. Science and Mathematics Education*. Dordrecht: Kluwer Academic.

Valdés, Eduardo Aldana. 1996. "Medición del logro educativo: el caso colombiano." PREAL Seminar on Evaluation and Standards in Latin American Education, Rio de Janeiro.

Valverde, Gilbert A. 1996. "Las políticas educacionales nacionales y la participación en estudios internacionales." PREAL Seminar on Evaluation and Standards in Latin American Education, Rio de Janeiro.

Vianna, Heraldo Marelím. 1996. "Programas de avaliação de sistema de educação estadual do estado de Sao Paulo." PREAL Seminar on Evaluation and Standards in Latin American Education, Rio de Janeiro.