



# **EDUCATIONAL ASSESSMENT SYSTEMS IN LATIN AMERICA**

Current Practice and Future Challenges

Guillermo Ferrer



# **Educational Assessment Systems in Latin America:**

## **Current Practice and Future Challenges**

---

Guillermo Ferrer



©2006 Partnership for Educational Revitalization in the Americas (PREAL).  
All rights reserved.

For additional copies, contact PREAL at the Inter-American Dialogue. The book  
can be downloaded from [www.preal.org/publicacion.asp](http://www.preal.org/publicacion.asp).

Inter-American Dialogue  
1211 Connecticut Avenue, NW, Suite 510  
Washington, DC 20036  
202-822-9002  
[iad@thedialogue.org](mailto:iad@thedialogue.org)  
[www.thedialogue.org](http://www.thedialogue.org) and [www.preal.org](http://www.preal.org)

Corporation for Development Research  
Santa Magdalena 75, Piso 10, Oficina 1002  
Santiago, Chile  
56-2-334-4302  
[infopreal@preal.org](mailto:infopreal@preal.org)  
[www.preal.org](http://www.preal.org)

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

ISBN: 0-9772271-4-6  
First edition  
Published in the USA

Book design: Nita Congress  
Cover design: Tim Cain

---

# Sponsoring Organizations

PREAL was established by the Inter-American Dialogue in Washington, D.C., and the Corporation for Development Research 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 nongovernmental 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 regionwide 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 through the generous support of the U.S. Agency for International Development, the Inter-American Development Bank, the Tinker Foundation, the International Association for the Evaluation of Educational Achievement, the World Bank, the GE Foundation, and others.

The Inter-American Dialogue is the leading U.S. center for policy analysis, exchange, and communication on issues in Western Hemisphere affairs. The Dialogue brings together public and private leaders from across the Americas to address key hemispheric problems and opportunities. Its

**PREAL**

**Inter-  
American  
Dialogue**

select membership of 100 distinguished citizens from throughout the Americas includes political, business, academic, media, and other nongovernmental leaders. Eleven Dialogue members have served as presidents of their countries, and nearly 30 have served at the cabinet level.

Dialogue activities are directed at generating new policy ideas and practical proposals for action, and getting these ideas and proposals to government and private decision-makers. The Dialogue also offers diverse Latin American and Caribbean voices access to U.S. policy debates and discussions. Based in Washington, D.C., the Dialogue conducts its work throughout the hemisphere. A majority of our Board of Directors are from Latin American and Caribbean nations, as are more than half of the Dialogue's members and participants in our other leadership networks and task forces.

Since 1982—through successive Republican and Democratic administrations and many changes of leadership elsewhere in the hemisphere—the Dialogue has helped shape the agenda of issues and choices in inter-American relations.

## **CINDE**

The Corporation for Development Research (CINDE) is a private, nonprofit institution based in Santiago, Chile. Founded in 1968, CINDE provides a nonpartisan academic environment for interdisciplinary research on national and international development issues. CINDE is a decentralized organization supported by a broad network of outside contributors. It sponsors research projects, seminars, workshops, and working groups whose findings may be freely published. CINDE provides a forum for professionals and specialists from various countries and cultural and professional backgrounds to meet, exchange information, and debate.

---

# Contents

|  |            |
|--|------------|
| <b>Acknowledgments .....</b>   | <b>vii</b> |
| <b>Acronyms and Abbreviations .....</b>  | <b>ix</b>  |
| <b>Introduction .....</b>  | <b>1</b>   |
| <b>Part I. Overview of National Educational<br/>Assessment Systems .....</b>       | <b>7</b>   |
| Considerations in Designing Good Assessment<br>Systems .....                       | 9          |
| Status and Characteristics of National<br>Assessment Systems in Latin America..... | 15         |
| Conclusions .....  | 46         |
| <b>Part II. National/Subnational Snapshots.....</b>                                | <b>53</b>  |
| Argentina .....  | 55         |
| Bolivia .....  | 60         |
| Brazil.....  | 64         |
| Chile.....   | 69         |
| Colombia.....  | 75         |
| Costa Rica .....   | 80         |
| Cuba.....  | 84         |
| Dominican Republic .....   | 87         |
| Ecuador.....   | 91         |
| El Salvador .....  | 94         |
| Guatemala.....   | 99         |
| Honduras .....   | 102        |
| Mexico .....   | 106        |

|                             |            |
|-----------------------------|------------|
| Nicaragua .....             | 111        |
| Panama .....                | 113        |
| Paraguay .....              | 117        |
| Peru.....                   | 121        |
| Uruguay.....                | 124        |
| Venezuela .....             | 129        |
| Minas Gerais, Brazil .....  | 132        |
| Paraná, Brazil .....        | 136        |
| São Paulo, Brazil .....     | 140        |
| Bogotá, Colombia .....      | 144        |
| Aguascalientes, Mexico..... | 148        |
| <b>References.....</b>      | <b>153</b> |
| <b>Interviews .....</b>     | <b>155</b> |



---

# Acknowledgments

This book draws substantially on information and insight provided by public officials and experts working on educational assessment throughout the region. Thanks also go to Andrew Crawley for translation, Nita Congress for editing and layout, Timothy Cain for cover design, and PREAL staff member Kristin Saucier for managing the production and publication process.

The book could not have been produced without the generous support of the U.S. Agency for International Development (USAID), the International Association for the Evaluation of Educational Achievement (IEA), the World Bank, the Inter-American Development Bank (IDB), the Tinker Foundation, the GE Foundation, and others. Their sustained support for education reform in Latin America has been crucial to developing the institutions and networks that have made this project possible.



---

# Acronyms and Abbreviations

|         |   |
|---------|---|
| ALL     | Adult Literacy and Lifeskills   |
| ANEP    | Administración Nacional de la Educación Pública (National Administration for Public Education – Uruguay)  |
| APRENDE | Segundo Proyecto de Educación Básica (Second Project for Basic Education – Nicaragua)   |
| APRENDO | Sistema Nacional de Medición de Logros Académicos (National System for Measuring Academic Achievement – Ecuador)  |
| CBC     | Contenidos Básicos Comunes (Common Basic Contents – Argentina)  |
| CECE    | Centro para el Estudio de la Calidad Educativa (Center for the Study of Educational Quality – Panama)   |
| CENP    | Coordenadoria de Estudos e Normas Pedagógicas (Coordinating Office of Teaching Studies and Norms – São Paulo, Brazil)   |
| DIGEBI  | Dirección General de Educación Bilingüe Intercultural (General Directorate of Bilingual Intercultural Education – Guatemala)                                    |
| DiNIECE | Dirección Nacional de Información y Evaluación de la Calidad Educativa (National Directorate for Information and Assessment of Educational Quality – Argentina) |
| ENEM    | Examen Nacional do Ensino Médio (National High School Examination – Brazil)   |
| ETS     | Educational Testing Service   |
| EVEP    | Evaluación de la Educación Primaria (Primary Education Evaluation – Mexico)   |
| EXANI   | Examen Nacional de Ingreso a la Educación Media Superior (National University Entrance Exam – Mexico)   |
| FDE     | Fundação para o Desenvolvimento da Educação (Foundation for Educational Development – São Paulo, Brazil)  |
| ICFES   | Instituto Colombiano para el Fomento de la Educación Superior (Institute for the Development of Higher Education – Colombia)                                    |

|         |   |
|---------|---|
| IDANIS  | Instrumento de Diagnóstico de Alumnos de Nuevo Ingreso a Secundaria (Instrument for Testing New Secondary School Pupils – Mexico)                 |
| IALS    | International Adult Literacy Survey   |
| IEA     | International Association for the Evaluation of Educational Achievement   |
| IIMEC   | Instituto de Investigaciones para el Mejoramiento de la Educación Costarricense (Research Institute for the Improvement of Costa Rican Education) |
| INEE    | Instituto Nacional de Evaluación de la Educación (National Institute for Educational Assessment – Mexico)   |
| INEP    | Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (National Institute of Educational Studies and Research – Brazil)          |
| LLECE   | Laboratorio Latinamericano de Evaluación de la Calidad de la Educación (Latin American Laboratory for the Evaluation of Educational Quality)      |
| MEC     | Ministério da Educação (Ministry of Education – Brazil)   |
| OAS     | Organization of American States   |
| OECD    | Organisation for Economic Co-operation and Development  |
| OREALC  | Oficina Regional de Educación para América Latina y el Caribe (Regional Office for Latin America and the Caribbean – UNESCO)                      |
| PAES    | Prueba de Aptitudes para Egresados de Educación Media (Learning and Aptitude Test for High School Students – El Salvador)                         |
| PEEP    | Proyecto Eficiencia de la Educación Primaria (Primary Education Enhancement Program)  |
| PER     | Prueba de Evaluación de Rendimiento Escolar (School Achievement Test – Chile)   |
| PIRLS   | Progress in International Reading Literacy Study  |
| PISA    | Programme for International Student Assessment  |
| PREAL   | Partnership for Educational Revitalization in the Americas  |
| PRONERE | Programa Nacional de Evaluación del Rendimiento Escolar (National Program for School Achievement Assessment – Guatemala)                          |
| SABE    | Solidificación del Alcance en la Educación Básica (Strengthening of Achievement in Basic Education – El Salvador)                                 |
| SABER   | Sistema Nacional de Evaluación de la Calidad de la Educación (National System for Evaluating Educational Quality – Colombia)                      |
| SACMEQ  | Southern and Eastern African Consortium for Monitoring Educational Quality  |

---

|         |   |
|---------|---|
| SAEB    | Sistema Nacional de Avaliação da Educação Básica (Basic Education Assessment System – Brazil)   |
| SARESP  | Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo (System for the Assessment of School Performance in São Paulo)          |
| SECE    | Sistema de Evaluación de la Calidad de la Educación (System for the Assessment of Educational Quality – Cuba)                             |
| SEP     | Secretaría de Educación Pública (Public Education Secretariat – various countries)  |
| SIMAVE  | Sistema Mineiro de Avaliação da Educação Pública (Educational Evaluation System – Minas Gerais, Brazil)                                   |
| SIMCE   | Sistema de Medición de la Calidad de la Educación (National System for the Assessment of Educational Quality – Chile)                     |
| SIMECAL | Sistema de Medición y Evaluación de la Calidad de la Educación (System for Measuring and Evaluating the Quality of Education – Bolivia)   |
| SINEA   | Sistema Nacional de Evaluación de los Aprendizajes (National System for Measuring Learning – El Salvador)                                 |
| SINEA   | Sistema Nacional de Medición y Evaluación de Aprendizaje (National System for the Measurement and Assessment of Learning – Venezuela)     |
| SINECE  | Sistema Nacional de Evaluación de la Calidad de la Educación (National System for Assessing the Quality of Education – Panama)            |
| SINMELA | Sistema Nacional de Medición del Logro Académico (National System for Measuring Academic Achievement – Guatemala)                         |
| SNE     | Sistema Nacional de Evaluación de la Educación Básica y Media (National System for Assessing Basic and High School Education – Nicaragua) |
| SNEPE   | Sistema Nacional de Evaluación del Proceso Educativo (National System of Evaluation of the Educational Process – Paraguay)                |
| TIMSS   | Trends in Mathematics and Science Study   |
| TIMSS-R | Trends in Mathematics and Science Study-Repeat  |
| UMC     | Unidad de Medición de Calidad (Quality Measuring Unit – Peru)   |
| UMCE    | Unidad Externa de Medición de la Calidad de la Educación (Unit for Measuring Educational Quality – Honduras)                              |
| UMRE    | Unidad de Medición de Resultados Educativos (Educational Results Measuring Unit – Uruguay)  |
| UNESCO  | United Nations Educational, Scientific and Cultural Organization  |
| USAID   | U.S. Agency for International Development   |



---

# Introduction

Since 2002, the Working Group on Assessment and Standards of the Partnership for Educational Revitalization in the Americas (PREAL) has been conducting studies on how the quality of education in Latin America is measured and assessed, as well as about the progress countries in the region have made in proposing and designing clear expectations of learning—which are, in this book, termed “standards.” Three of those studies establish the conceptual basis of this report on the current status of national assessment systems and of standards development in the region:

- “¿Cómo presentan sus resultados los sistemas nacionales de evaluación educativa en América Latina?” (Ravela 2001a);
- “Los próximos pasos: ¿Hacia dónde y cómo avanzar en la evaluación de aprendizajes en América Latina?” (Ravela 2001b); and
- *Aspectos del curriculum prescrito en América Latina: Revisión de tendencias contemporáneas en curriculum, indicadores de logro, estándares y otros instrumentos* (Ferrer et al. 1999).

This book moves these studies a step further in documenting the progress Latin American and Caribbean countries have made in establishing, consolidating, improving, and using national and subnational systems to assess learning and in devising academic achievement standards for basic education.

Initial research efforts were geared to organizing the available information on national assessment systems, par-

ticularly about agencies responsible for testing in each country, the types of instruments devised, the student populations tested, and the use made of the information provided by the systems. Data compiled in recent years by other researchers (Rojas and Esquivel 1998, and Wolff 1998) were collected, updated, and expanded through interviews and documentary analysis conducted in almost all countries of the region.

## Organization

*Part I* discusses progress made by Latin American countries in their efforts to establish clear expectations of learning, devise appropriate measurement and assessment tools that yield information on the extent to which such expectations are actually being met, inform educational actors and the general public of the quality of public education systems in terms of achieved learning, and devise and adjust education policies on the basis of the data provided by the assessments so as to improve learning.

The first section of part I summarizes the processes and conditions that, according to the Working Group on Assessment and Standards and other studies, would help consolidate a robust and legitimized assessment culture in Latin American countries and have a greater impact on policy-making and teaching practices in order to improve student academic achievement.

The following section provides an extended description of each of the analytical categories chosen for this review of national systems, and offers a narrative summary of some national and subnational cases that illustrate the different technical and organizational arrangements for assessment that have been developed in the region. Here and throughout the book, the most institutionalized or most innovative cases are presented, although occasional reference is made to national systems that serve as negative examples in order to illustrate the types of problems that arise in each aspect of assessment analyzed here. Part I concludes with a summary discussion of the most notable advances and challenges apparent in the region's assessment systems.

*Part II* provides individual summaries, or snapshots, of each national and subnational case of educational assessment, using the same analytical categories described and



discussed in part I. Readers may refer to these individual summaries for more detailed information on the cases mentioned throughout the text, or for cross-country comparisons.

While the study was being prepared, national assessment systems throughout Latin America underwent institutional and methodological changes, often coinciding with the frequent shifts in political authority in the education sector. In some cases, these changes amounted to partial adjustments in the methodology or coverage of tests; in others, there has been substantial structural and institutional change. An effort has been made to provide the most recent information available at the time of writing, although some recent changes might not be reflected herein. Readers should regard this analysis as a dynamic portrait that primarily covers the development of national assessment systems throughout the 1990s, as well as some of the most significant changes recently introduced in response to problems faced by the various countries.

Obviously, the description offered here could be updated daily, especially given the institutional instability besetting many of the region's assessment systems and education ministries. Nonetheless, it is hoped that the study fulfills its prime function of showing the main trends, advances, and problems that have become apparent in the past 10 to 15 years of standardized testing.

The information presented in this book is based on bibliographic and documentary sources, and on almost 60 interviews with officers of assessment units and experts in the field in 19 countries and 5 subnational systems. Additional information was collected via electronic communications with those officers and experts, who supplemented, modified, or updated the data sent them for review.

Each issue or category of analysis considered here was broken down into a set of themes, as shown in the following table.

Since this is a critical analysis of national assessment systems rather than a description of them, the author's personal opinions of system features, of how they have developed in

## Methodology

recent years, and of their impact are presented. Nonetheless, the assertions made here are based on an analysis of all available sources covering each national and subnational case, and efforts have been made to include the opinions of various actors in the assessment process, both from within and outside the official management structures.

It was also necessary to compare the perspectives of government actors from different periods; these often corresponded to national governments with different political inclinations and management styles. This multiplicity of voices has necessitated some amount of summarization and interpretation—which doubtless could be a matter of debate in each case examined here. It should be noted, however, that all the cases analyzed were reviewed by technical authorities in each country, each of whom had an opportunity to correct, moderate, or update the information and interpretations offered by the author.

| Category of analysis     | Themes   |
|--------------------------|--|
| Institutional framework  | <p>Adequacy to political framework and technical capabilities:</p> <ul style="list-style-type: none"> <li>• Stability</li> <li>• Financing and administrative autonomy</li> <li>• Human resources</li> <li>• Autonomy and capacity to disseminate results</li> <li>• Transparency</li> </ul>   |
| Curriculum and standards | <ul style="list-style-type: none"> <li>• Availability, adequacy and use of national curricular frameworks to design assessment instruments</li> <li>• Development and validation of standards consistent with the prevailing curriculum</li> </ul>   |
| Instruments              | <ul style="list-style-type: none"> <li>• Solid and explicit conceptual framework for drawing up reference matrices</li> <li>• Validation of reference matrices and instruments</li> <li>• Types of items</li> <li>• Cultural and linguistic adequacy (especially where bilingual and intercultural education programs should or do exist)</li> <li>• Sample-based or census-based coverage consistent with the aims of the assessment</li> <li>• Study of the in-school and out-of-school context for analysis of performance-related factors</li> </ul> |
| Reports                  | <ul style="list-style-type: none"> <li>• Coherence between types of reports and expected uses (curricular development, pedagogy, targeting support, teacher training, selection of students, and so on)</li> <li>• Adequacy of reports for different audiences (clarity, guides to interpretation, sensitization, and so on)</li> <li>• Information on associated factors and value-added models</li> </ul>  |
| Dissemination and uses   | <ul style="list-style-type: none"> <li>• Delivery: time frames; scope; regularity</li> <li>• Impact: school use; policymaking/program design; political accountability</li> <li>• High stakes: schools; teachers; students</li> </ul>  |
| Subnational systems      | <p>Main differences from national systems:</p> <ul style="list-style-type: none"> <li>• Standards</li> <li>• Sampling (or censal) coverage</li> <li>• Participation of local actors</li> <li>• Time frames and formats for delivering results</li> <li>• Use for pedagogical improvement</li> </ul>  |
| International tests      | <ul style="list-style-type: none"> <li>• Development of technical capacities</li> <li>• Dissemination of results and impact on public opinion</li> <li>• Specific uses of the results</li> </ul>   |



**Part I.**

---

# **Overview of National Educational Assessment Systems**



## Considerations in Designing Good Assessment Systems

The models and systems<sup>1</sup> designed for measuring and assessing student academic performance obviously reflect a country's political conditions and technical and operational capacities, both current and potential. As evidenced by the Latin American experience of educational assessment thus far, as well as experiences in more developed countries, certain considerations should be taken into account by a country that is planning or adjusting its system of measuring and assessing learning. The most significant of these considerations are outlined below.

The purposes of educational assessments and the uses to which the results will be put are issues that should be subject to extensive debate and consideration from the moment that a testing program or more ambitious evaluation system is instituted. The advantages of particular approaches, methodologies, and instruments, as well as the expected impact of the assessments, should be determined according to how well they serve those purposes and how effectively the results can be used in meeting those purposes.

Countries must consider what is the most appropriate institutional framework for the evaluation agency. It is especially helpful to discuss and clearly define—in line with each country's political and technical circumstances—if the assessment agency should be an official body, whether external experts should be contracted, or whether a mixed system should be adopted. Whatever the case, it is crucial that the assessment agency's degree of functional autonomy be defined clearly.

### Goals and Purpose

### Institutional Framework

---

<sup>1</sup>The term “system” denotes relatively complex organizations and the linkages among their components. Although the term is here used to refer to all national education assessment programs, note that in some cases, these “systems” are not linked to other policy endeavors in their respective education sector or are only linked at certain points or periods; they are thus more aptly regarded as assessment “projects” rather than systems proper.

### **Relationship to Curricular Goals**

The institutional framework can have significant implications for the credibility and legitimacy of assessment activities; the technical and operational sustainability of the assessment system; and the potential dissemination, use, and impact of assessment results.

The assessment instruments must be consistent with the educational expectations or goals a country has adopted through its national curriculum. These goals must be expressed with sufficient clarity so that the cognitive and disciplinary skills attained by the students can be identified as objectively as possible by all the evaluators in the system: teachers, principals, supervisors, and professionals from outside the school.

These goals should be described in operational terms,<sup>2</sup> and examples should be provided, so all actors in the system have a precise understanding of the expectations for academic achievement in the various areas of the curriculum. Moreover, the description should distinguish between, and give examples of, the different performance levels that can be attained in each skill in the various curricular areas under assessment. For example, it could specify an inadequate level of performance, as well as the minimum, average, and advanced levels, for each curricular area and each grade level being assessed. This classification, together with its operational description and examples, comprises the basic interpretative framework that allows education professionals and the general public to understand the significance of the data provided by the assessments. This framework can also give rise to targeted pedagogical and curricular strategies that may improve educational outcomes.

When the complex skills to be developed are not clearly defined in operational terms, and when there are only some

---

<sup>2</sup>An operational definition of learning targets consists of a clear and sufficiently detailed description of the kind of conceptual and procedural knowledge that students must prove they have acquired. This does not imply such a degree of prescriptiveness that observation and evaluative judgment are limited to a single form of expression of knowledge, but it should be sufficiently descriptive to guide the development of valid assessment instruments, ensure uniform criteria in grading tests, and produce comprehensible and substantive information on the kind of knowledge students have acquired.



general guidelines for curricular development (as is the case in almost all Latin American countries), the linkages and communication between the assessment agencies and the professionals responsible for designing the official curriculum have to be strengthened. The aim should be joint development of clear indicators for academic performance and of assessment instruments aligned with those indicators.

The methodology and coverage of the assessments must be consistent with their purposes and uses; to this end, they should be validated socially, as well as technically.<sup>3</sup> Decision-makers should carefully determine the aim and subjects of the assessments—that is, if their goal is to assess the system, the institutions, or the individual actors (especially the teachers and students). Decisions on these matters will determine, for example, if the tests will be census-based or sample-based, if they will be norm-referenced or criterion-referenced, and if the results will hold high or low consequences (“stakes”) for the actors.

The assessment instruments must be specifically tailored to meet the agreed-upon purpose. If the data provided by the assessments are to be used to improve teaching or the curriculum, the instruments must reflect this. That will require decisions on, for example, the evaluation model used to assess the achievement of increasingly complex skills; whether the test items call for multiple-choice answers, open answers, practical laboratory demonstrations, or a combination of these options or others that might be proposed; and the extent to which the test will cover the curriculum.

All assessment instruments must be experimentally validated through pilot field applications before the final versions are drawn up. Some countries will need specific validations when a decision is made to assess groups whose native language is not the official language of the country. In such cases, the instruments’ technical validation will entail both

## **Assessment Instruments**

---

<sup>3</sup>Social validation refers to a process whereby education community stakeholders—including civil society representatives—are engaged in discussions and the decisionmaking process to ensure the appropriateness, relevance, and quality of assessment content and instruments.

linguistic considerations (either appropriate translations or construction of test items in native languages) and general cultural considerations to ensure that items are not biased.

If the assessment instruments will also be used to assess the in-school or out-of-school factors related to student academic performance (that is, to measure variables that could determine or be associated with performance and to analyze their links to it), technically robust procedures must be used to gather and statistically analyze the data. Standardized instruments, such as interviews, observations, analysis of classroom video recordings, etc., can also be devised for systematic qualitative follow-up on teaching methods and institutional processes that might help explain the different results yielded by quantitative assessments.

## **Results Reporting**

The way in which test results are reported should be consistent with the goals of the assessment and should take into account users' specific information needs. For example:

- If the aim is to provide information on student achievement of certain subjects in recently implemented areas of the curriculum so that schools can make adjustments where they detect difficulties, it is of little use to report results in aggregate form with a gross percentage of achievement for the whole of each curricular area or subject matter.
- If the aim is to encourage parents to become more involved in their children's education, they have to be given information not only on the level of performance their children have achieved in certain academic capacities but also a description and examples of the kind of performance that is expected at their grade level.
- If the aim is to make judgments about the quality of schools in terms of students' academic achievement—and not to use the results to compare or rank institutions or geographic areas—results must be reported so as to distinguish the economic and socio-cultural levels of the students that the schools serve, as well as the schools' infrastructure, facilities, and

human resources. It is thus possible to compare each school's level of achievement with that of all schools operating under the same conditions, thereby obviating the risk that unfair conclusions might be drawn about the quality of each institution.<sup>4</sup>

- If the assessments follow a “value-added” model—that is, appraising students' academic progress relative to themselves over time, and/or relative to students of a similar socioeconomic level—data can be provided and fairer inferences drawn about each school's capacity to offer adequate educational opportunities.

Assessments geared to gathering information that allows remedial intervention in order to improve outcomes should be accompanied by solid programs, targeted or wide ranging, that help improve management and teaching practices in the schools.

In the final analysis, assessment systems will only be consolidated and find social legitimacy when the stakeholders regard them as a tool to improve educational quality, and not as a purely diagnostic or even punitive exercise.

If assessment efforts are geared to mobilizing public opinion and social pressure for educational improvement, good communication strategies are vital. The effectiveness of such strategies depends not only on a system's capacity to reach different audiences through mass distribution of publications, but also on the degree to which key actors and organizations of the educational community and civil society are involved in the assessment process. To the extent that those actors—such as representatives of teachers' unions, entrepreneurs, or the media—take part in these processes and in designing communication strategies, the information provided by the evaluations will be broader in scope and have a greater impact. Broad consultation with these stakeholders is necessary even if results are only to be disseminated locally

## **Dissemination and Use of Results**

---

<sup>4</sup>For a more detailed discussion and examples of this aspect of measuring learning, see Ravela (2001a).

or in a highly targeted manner (i.e., to individual schools or municipalities) in order to determine most precisely the best communication strategy to use.

Regardless of the main intended audience, the information must be timely and delivered on a regular basis, and must reach the greatest possible number of individuals and institutions. Assessment efforts in each country require central (national) political will and initiative if they are to be instituted and consolidated and have a significant impact. It is also helpful, however, to facilitate and foster subnational initiatives (in specific regions and jurisdictions) in the area of external assessment. Cooperation between central and local technical/political agencies is crucial in developing countries to create forums for experimenting with assessment practices, institutionalizing them, and ensuring that they have greater impact.

### **Participation in International Tests**

If a country decides to take part in comparative international tests of academic achievement, then—just as with national tests—the country’s specific goals must be defined clearly, so that communications strategies and interventions can be devised to strengthen the impact of the information provided by the assessment.<sup>5</sup>

---

<sup>5</sup>Ferrer and Arregui (2003) address this issue in depth.

## **Status and Characteristics of National Assessment Systems in Latin America**

In the past two decades, most Latin American countries have adopted some model for national educational assessment. In some cases, assessment experience has been slowly acquired through the administration of tests for highly specific purposes, such as regulating admission to higher education or gauging the impact of a targeted educational improvement project financed by an international organization. In other cases, assessment through standardized tests was adopted quite abruptly as part of the more or less comprehensive education reform programs common in the 1990s, and is conducted within institutional and organizational frameworks created specifically for assessment purposes. There is wide variation in the level of development of these national systems, depending on the technical, operational, and financial capacities of the individual assessment agencies and the political contexts in which they operate.

This section builds on the foregoing discussion of the desirable characteristics of national assessment systems as outlined in the literature and summarized above, and offers examples drawn from national and subnational cases of educational assessment gathered in the course of the research for this study. For clarity and consistency, this discussion of system characteristics is organized under the same headings as in the previous section. At the end of the section, there is a brief summary of Latin American countries' participation in comparative international learning tests; this information is based on data collected in previous research by the author.

In many Latin American countries, insufficient thought has been given to the specific purposes assessment systems are expected or desired to serve; this criticism has been leveled since PREAL's Working Group on Assessment and Standards began its activities in 1998.<sup>6</sup> As a result, some of the techni-

### **Goals and Purpose**

---

<sup>6</sup>See the introduction to Ravela (2001b).

cal definitions initially adopted were not the most appropriate for the various purposes and uses the assessment instruments and results were intended to have. It is difficult for a single design to meet all the (sometimes mutually inconsistent) objectives policymakers pursue over time, but the cost of designing complementary instruments and the urgent need for information to underpin certain decisions can sometimes give rise to the erroneous use of test results. That circumstance is exacerbated by an absence in many cases of a long-term vision of how the system should develop, and the need to improvise “emergency” measures that sometimes undermine systems’ legitimacy and hinder their consolidation.

Fortunately, the accumulation of experience and the relative strengthening of assessment capacities in many of the region’s countries now allow those goals to be reviewed: the debate can today be enriched by actual experience, and it need not rely solely on theoretical concepts or models imported from elsewhere. As is evident in the Dominican Republic, Chile, and other countries, it is possible to make substantial adjustments without entirely dismantling existing systems or established institutional capacity.

## **Institutional Framework**

Latin America has at least two decades of experience with educational assessment, although most of the official systems were established in the 1990s. Some countries, such as Chile and Colombia, have been continuously administering tests of academic achievement for many years, and their assessment agencies enjoy a high degree of legitimacy and continuity despite the political vagaries periodically prompted by changes in national government. Other systems, such as those in Bolivia, Ecuador, Guatemala, and Peru, have been more vulnerable to the political will of successive governments. As a result, the information that these systems produce is not always made public, and sometimes technical staff are abruptly changed. These circumstances affect the quality of the assessment instruments and exercises, and the design and implementation of appropriate strategies to disseminate results.

Countries have tried various institutional frameworks for managing their assessment systems, and have learned and

progressed much by searching for the best means to accommodate the interests and goals of state education policies. In general, despite changes in governments and goals—and even, in some cases, during times of severe political instability—national assessment systems have tended to survive and have been reestablished or reorganized when their activities have been suspended.

In several countries, the most stable institutional arrangements are those that were established outside the organizational structure of the education ministries. Those systems, generally referred to as assessment “institutes,” have greater administrative and technical autonomy than other line agencies. They typically conduct assessments and report results with greater flexibility and consistency than systems that are dependent on the ministries.

Two examples of this kind of institutional arrangement are the Colombian Institute for the Development of Higher Education (ICFES) and Brazil’s National Institute of Educational Studies and Research (INEP). Although Colombia’s Ministry of Education has an assessment unit, ICFES has taken the technical and operational lead in the country’s various national assessments. ICFES is not wholly independent of the ministry, which finances and presides over it, but it has wide discretion and autonomy in making technical decisions on some of the national tests, such as the state examination for admission to higher education. In Brazil, national assessments are undertaken by INEP, which is managed independently of the Ministry of Education. In both countries, regular assessment exercises have been conducted for several years: the technical teams have had a chance to establish themselves and have gradually developed new assessment technologies.

Mexico, which had a long-standing assessment system dependent on the country’s Secretariat of Public Education, recently created the National Institute for Educational Assessment. This semiautonomous agency has greater latitude in making technical and administrative decisions on assessment exercises than did the former General Directorate for Education. In 2001, Argentina also tried to establish a semiautonomous assessment institute, the Institute for

Educational Quality, but unexpected changes in the national government that year led to the institute's dissolution and the reinstallation of the national system; this system is now called the National Directorate for Information and Assessment of Educational Quality and is part of the Ministry of Education.

Autonomous assessment entities may have a higher degree of functional independence and technical legitimacy than those associated with ministries, but they can also create problems. The greatest risk is that they might become disconnected from ministry information needs and turn into programs that, while of a high technical quality, have little impact on policy decisions geared to improving educational quality. It will be possible to evaluate in the coming years how the recently established national institutes have worked out and what impact they have had.

In other countries, the trend has been reversed, with systems or projects that were previously administered by independent institutions now managed by national ministries of education. This is the case in Chile, where the Ministry of Education originally delegated the organization and implementation of the National System for the Assessment of Educational Quality (SIMCE) to public or private universities that had the requisite technical capacity to carry out those tasks. Today, however, the assessment system is managed entirely by the Ministry of Education's curriculum and assessment unit. In this case, given the technical capacity and legitimacy that SIMCE has acquired over the past two decades, locating the system inside the Ministry of Education does not seem to have diluted its independence or undermined its stability.<sup>7</sup>

The transfer of responsibilities to education ministries has not always been either successful or beneficial. In Ecuador, termination of the National System for Measuring Academic Achievement (APRENDO) assessment program, which

---

<sup>7</sup>SIMCE's hard-won legitimacy is due to the fact that test results have shown signs of substantial improvement in the quality of education during the past decade. From a political viewpoint, SIMCE is a critical tool for ensuring the outcomes of the national government's education policies. Whether that legitimacy would be affected if results improved and SIMCE were seen as a means of "validating" ongoing education policies remains to be seen.



was financed completely by international aid, meant the end of the assessment exercises. Plans were made to transfer the system to public administration, but the ministry's lack of material resources and specialized professionals made it impossible to continue the assessments as before; no new programs have since been devised.

This phenomenon is in fact common to almost all countries whose assessment systems were established with financing and technical support from international organizations, mainly that of the Inter-American Development Bank and the World Bank. There are some exceptions, such as Uruguay and Argentina, which now finance their systems with resources from the ordinary budget within the ministerial structure; the continuance of their tests has not been significantly affected.<sup>8</sup>

Another factor that seems to help make assessment systems stable and permanent is the existence of national laws that call for assessment systems to be created and maintained. An example is the system of national tests in the Dominican Republic. Although the system operates within somewhat unsettled conditions, the annual assessment exercises are conducted in compliance with the general education law and a series of specific national ordinances. The same could happen in Peru, where a new general education law provides for the creation of the Peruvian Institute for Educational Assessment, Accreditation, and Certification, and for permanent support to it.

The subnational assessment systems analyzed in this study are administered entirely by the central or district-level education secretariats, an institutional arrangement that seems to work satisfactorily. Only the System for the Assessment of School Performance in São Paulo (SARESP) has a different institutional structure, since an agency independent of the government takes the lead in technical matters. Perhaps because these education systems are smaller and more easily managed, they feature a higher level of communication and coordination with other offices in the sector. This in turn

---

<sup>8</sup>The exact number of countries that ended their assessment programs because of external and internal financing problems has yet to be determined.

facilitates information exchange and increases the possibility that the system will have an impact on education policymaking—which is not the case in centrally administered systems. Other successful subnational systems encourage greater participation by local actors, especially teachers, parents, and academic organizations involved in teacher training.<sup>9</sup>

### **Relationship to Curricular Goals**

One of the greatest concerns expressed by those responsible for designing assessment instruments is that national curricula lack clear (or even operational) definitions of what students are expected to be able to do with the conceptual knowledge contained in the curricula. Almost always, the curricular matrix for developing assessment instruments is devised by means of a “specifications table” that outlines how test items should be drawn up so as to cover a given number of priority goals—although they do not account for all of the official curriculum or its complexity. Naturally, the negative outcome of this situation is a weak framework for interpreting results, since there is no professional or societal agreement as to what the students in the system should know and be able to do. Added to this is the problem that content that can be measured easily is often accorded priority, even if it is not necessarily the most important content as perceived by system stakeholders.

These problems are compounded by the often poor communication between assessment agencies and those responsible for designing and developing the national curriculum—a circumstance that stems in part from the different professional backgrounds of the respective personnel. Given the vagueness of official educational targets, this poor communication is normally one of the chief obstacles to ensuring that assessment activities are consistent with those targets and can facilitate their being met.

Some countries are moving toward establishing certain specifications, on which the evaluators themselves have reached agreement, as to what students are expected to learn, even though the country’s curricular development processes are not consistent with that course of action. Many countries

---

<sup>9</sup>Several of these subnational cases are described in part II.

in the region have professional assessment staff dedicated to establishing or specifying clear curricular targets that allow them to design more focused tests, while they continue to work (and sometimes further the debate) on what concrete outcomes of learning are expected as a priority from the students. Given the lack of concerted efforts to develop content and academic performance standards, the assessment agencies' endeavors are a significant step in the right direction. Colombia, Ecuador, and Uruguay illustrate this trend, as do the subnational assessment systems in Aguascalientes, Mexico, and Bogotá, Columbia.

ICFES, the national Ministry of Education, and the Bogotá Education Secretariat have all made significant efforts to define complex and operational abilities that facilitate better assessment and interpretation of results. In Colombia, where the official curriculum consists of general guidelines and performance specifications are somewhat vague, such efforts have moved forward on at least three fronts. First, the Ministry of Education's assessment teams have defined mathematics and language skills. While in principle these definitions should have guided the assessments of a comprehensive follow-up program to be undertaken each year between 1996 and 2005, in actuality this program was discontinued following a substantial restructuring of the Ministry of Education. Second, ICFES has defined a significant set of complex learning skills in several areas of the curriculum covered by the state examination. These skills, which facilitate educational assessment at distinct and clearly defined achievement levels, offer a more robust framework for interpreting results than that provided by assessments designed on the basis of a typical specifications table. Finally, the Ministry of Education has recently sought to establish a set of national standards in three curricular areas. These were presented to the educational community and general public in 2003. It is still too early to judge the validity and impact of these standards, but their mere presentation has triggered debate in professional circles and prompted opinions in the media.

Ecuador's experience has been similar to that of Colombia, albeit on a smaller scale. In view of the uncertainty about the curriculum that was to be made official in the mid-1990s,

APRENDO's technical staff decided to define a set of basic academic skills that later facilitated the design of criterion-referenced assessment instruments. These tests feature four items per skill, and at least three have to be answered correctly for that skill to be regarded as attained. In mid-2001, Ecuador's Ministry of Education was planning to use this definition of skills and assessment to set up a system of public outreach on what the students and system had attained, using a cut-off line to serve as an achievement horizon that can be raised year by year if national results improve. The assessments have since been discontinued, however, following the most recent change in educational authorities; no announcements have been made about the future use of these curricular matrices.

In Uruguay, where the long-standing national curriculum puts significant technical obstacles in the way of defining reference matrices, the process of drawing up and administering the tests has created an interesting forum for debate on the curriculum. Specifically, the assessment questions call for demonstration of knowledge and cognitive aptitudes that are relevant and desirable, but that are not explicit in the national curriculum. Validation of the test items, especially by teachers, has spurred substantial thinking about the established curriculum and the way it is implemented in the schools.

A set of skills consistent with national curricular guidelines was also established for the district-level assessments undertaken by Bogotá's education secretariat. These tests, like the national-level State Examination, provide an explicit conceptual framework and operational definitions for each of the assessed skills. As a result, different levels of student performance can be reported and illustrated more clearly than would be possible using the national curricular guidelines. Both technically and politically, the basic skills that have been proposed can be regarded as curricular standards, since they offer a means of effective communication among all the actors in the sector as to what students in the system are expected to learn.<sup>10</sup>

---

<sup>10</sup>For more details on curricular and assessment development in Bogotá, see Ferrer (2004).

Aguascalientes is another interesting case. Here, the assessment's reference matrix does not involve a curricular proposal that differs from the national curriculum, but the learning goals chosen for the assessment constitute a concerted effort to prioritize a series of topics that are deemed to be basic and that all students in the system should be able to grasp. Such prioritization is useful because proposals by the schools and education secretariat to improve education are based on assessment results and reflect the scope of that skill set. There have been similar experiences in the subnational systems of Minas Gerais, Paraná, and São Paulo, although in those cases the authorities developed their own curricular frameworks based on national parameters.

In most cases, it would be useful to provide illustrative examples of what students should know and what their academic performance should be at the end of certain levels of education. The standards would then not only stipulate the content of learning but would also, through selected examples of test answers, shed light on the types and levels of performance expected of students. This information would help guide teaching and the development of best learning assessment practices in schools. It would also stimulate participation on the part of system beneficiaries and encourage them to demand better results. Initial steps in this direction are being taken in Argentina, El Salvador, Peru, and several other countries in the region.

Chile seems to be moving clearly and forcefully in this direction. With technical assistance from Australia, Chile's Ministry of Education is drawing up content and performance standards, as well as learning progress maps for students from first grade to the fourth year of high school in language, mathematics, history and the social sciences, the natural sciences (biology, chemistry, and physics), and English as a second language. The ministry has also arranged for standards to be established for early childhood learning. Using these instruments, the national tests should be able to report performance scores referenced to educational achievement goals that are clearly established before students take the tests and, of course, in sufficient time to influence teaching. Thus far, SIMCE has been unable to report results accord-

ing to clear criteria on the acceptable or satisfactory level of achievement or on the number of students who reach that level. If the assessment were linked to expected achievement, scoring would be more meaningful, and teachers, parents, and students would understand just how much a student knows and what the teacher should teach in order to meet the established goals. The first assessment in Chile based on learning standards will be undertaken in 2006.

## Assessment Instruments

**General Coverage.** Certain generalities exist regarding coverage and format of national assessment instruments. Most national assessment systems test learning in mathematics, language and communication, and the social and natural sciences. Some also assess achievement in foreign languages, especially English and French. In multilingual countries such as Bolivia, Guatemala, and Peru, tests are translated into and administered in indigenous languages. It is common in high school to test biology, physics, and chemistry separately; in the social sciences, there are different tests for history, geography, and civics.<sup>11</sup>

Almost all countries with national assessment systems measure achievement in both primary and secondary education. Tests are normally administered at the end of two- or three-year cycles. The reasoning behind this periodic coverage is that some subjects cannot be grasped fully by the end of a single academic year but require longer periods that coincide with age ranges and phases of cognitive development; this is termed multiyear curricular sequencing. Some countries, such as Chile, Costa Rica, and the Dominican Republic, administer tests when students leave middle or high school (some of these tests are called “baccalaureate examinations”), as a requirement of graduation or as a selection criterion for admission to higher education. Mexico also administers tests before students graduate from primary school to high school, although the results of these tests are used for diagnostic purposes and not for selection.

---

<sup>11</sup>Civics education, or education for civic responsibility and democracy, is becoming more prominent in national curricula, and some systems have devised specific instruments to assess learning in this area.

**Instrument Design.** One of the first steps in technically validating standardized assessment system design is to decide on the target population and curricular coverage of the assessment instruments so as to ensure that the data collected will be consistent with the uses to which the results will be put.<sup>12</sup>

Regarding population coverage, some countries opt to use *census-based tests*, while others use *sample-based tests*. Census-based tests are administered to all students at a certain stage of schooling and provide results by school and by student. They facilitate systems of incentives and penalties in line with achievement (high stakes), or make it possible to target support to those schools that have greater needs and face greater difficulty in the area of academic performance. Sample-based tests, on the other hand, seek to gather information on different groups in the system as a means of detecting general educational problems and, in principle, devising comprehensive support strategies for all the schools and students in a given social group or area of the country—specifically, those that need the most help.

Preferences regarding coverage and analysis of curricular achievement determine whether tests are *norm-referenced* or *criterion-referenced*. Norm-referenced tests seek to compare the achievements of different groups of students; the aim is not to provide a broader analysis of the kind of knowledge the student has, nor of its depth. Criterion-referenced tests, by contrast, assume that there is agreement on which basic skills or academic aptitudes all students should attain. The tests then establish a cut-off line or achievement score indicating minimum or satisfactory attainment, gauge whether students reach that level, and determine the extent to which they do so as a result of their schooling.

Both norm- and criterion-referenced tests can be sample- or census-based. Nonetheless, and as explained in great detail in previous studies by this working group (Ravela 2001b), wider coverage of the student population can limit the instruments' capacity to gather more detailed data on learning, and vice versa. Again, the decision on the kind of

---

<sup>12</sup>Ravela (2001b) provides a comprehensive analysis of this kind of validation for design purposes.

coverage to prioritize depends on the use to which the information will be put.

National systems that use criterion-referenced tests generally have more robust conceptual bases and offer sounder operational descriptions of learning and expected performance levels. Norm-referenced tests, by contrast, are often designed using a model that selects assessment questions in line with an index of discrimination (a measure that reflects differences in student achievement) revealed by the pilot tests; the questions do not necessarily reflect the relevance of the content assessed nor ensure that that content represents the entire reference curriculum.

In Uruguay, the tests are criterion-referenced. Test items are not chosen according to difficulty or a discrimination index but according to the “value” of the items—that is, those that reflect the kinds of skills the education system believes should be developed before students complete certain levels of schooling. Once the instruments have been designed, and following consultation with experts in curriculum design and teaching, a cut-off line is established; in Uruguay, this is 60 percent correct answers to all the questions.

The state examinations that ICFES designs in Colombia, which were originally norm-referenced, are now criterion-referenced. The change was initiated by a new team of technical professionals and by demands from students and the broader educational community that the tests focus less on simple rote learning and more on the development of complex cognitive capacities, while reflecting more up-to-date conceptual disciplinary knowledge. Updating the tests entailed aligning them more closely with official curricular guidelines, international standards of scientific literacy, and formative concepts that stress the need to educate involved citizens with a democratic outlook. The examinations assess skills in interpretation, argument, and assertion, which are judged to be “fundamental means of participating in and building the social sphere” (Asmar Amador et al. 2002).

Bogotá’s education secretariat sets criterion-referenced tests that are “quasi-census-based” (80 percent coverage of the grades being assessed). The criteria consist of a series of skills that are defined locally on the basis of national curricu-



lar guidelines. The results are reported as an average percentage of correct answers per type of skill—more specifically, by performance levels for each of them. Open answers, such as textual composition, are graded and reported according to preestablished competence levels that guide the teachers' grading and their interpretation of the results when they receive the school reports.

Several countries that began using norm-referenced tests—such as Argentina, Brazil, Chile, and Costa Rica—have opted to produce their results report in line with criteria defined later. This does not mean that the tests are strictly criterion-referenced, since the assessment model is normative, but criteria and cut-off lines are established for the report with a view to giving better information to system users. This strategy poses some risks for the validity of interpretations: unless the range of items is sufficiently wide, the items chosen according to their discrimination index might not represent attainment of the skills and knowledge being assessed.

The extent of the population covered by the tests varies from country to country. About nine countries use or have used only sample-based tests, two others have used only census-based tests, and eight have experimented with both. Tests in all the subnational systems analyzed in this study are census-based. In general, assessment coverage matches the uses to which the results are put: conclusions usually are not drawn on the basis of instruments that do not allow for such inferences. It is also true, however, that assessment data are alarmingly underused in designing strategies to improve educational quality; sometimes, too, they prompt unwarranted conclusions or invalid generalizations.

All countries use closed, multiple-choice questions; several also use open-ended questions calling for written responses, particularly when evaluating language and communication. Grading these open-ended questions has posed technical problems for some countries, in terms of defining or applying uniform assessment criteria. Countries are increasing their use of open-ended questions, despite their higher cost and the difficulties entailed in reviewing and organizing them, because they are more useful in identifying the procedures students use in trying to answer questions

and solve problems. As a result, they are more conducive to devising proposals on how to improve teaching and learning in some content areas and how to develop skills.

**Technical Validation.** Technical validation of test items or questions is a critical element in developing assessment instruments; unfortunately, national technical reporting on this subject is not particularly detailed, so it is difficult to provide much information about technical validation experiences in Latin America. International assessment theory provides guidelines and standards for validating test questions, but there is not enough information to determine the extent to which those standards are respected when national tests are developed in Latin America. Many technical reports state that teachers and experts in the field took part in the validation, but they do not indicate, for example, which specific criteria were used to judge the validity of the items (such as cultural biases, “noise” or the overlapping of content among disciplines, curricular relevance, and comparability across years). The quality of the instruments is an issue that merits more detailed study and will be addressed in future undertakings by the Working Group on Assessment and Standards.

**Language Adaptation.** Three countries in the region administer tests in indigenous languages. Peru administers tests in Quechua and Aymara; Bolivia in Quechua, Aymara, and Guarani; and Guatemala in the four main Mayan languages that are spoken by 84 percent of the country’s population. Guatemala ultimately had to abandon this effort, because a large proportion of Mayan-speaking students drop out of school.

**Factors Related to Academic Performance.** Almost all the assessment programs in the region normally complement their academic achievement tests with questionnaires for a variety of stakeholders in the educational community, especially students, teachers, principals, and parents. These questionnaires are intended to gather information on in-school and out-of-school factors that might be statistically correlated with student performance. In-school factors are normally indicators of the state of the infrastructure, teaching equipment, a school’s professional resources, student

attitudes toward the disciplines being assessed, teaching practices, and the extent to which the official curriculum is taught in the classroom.<sup>15</sup> Out-of-school factors are related in particular to students' economic and sociocultural levels (according to income information provided by the families), level of schooling of parents and siblings, household access to electricity and potable water and to cultural and information resources such as television, radio, books, newspapers, and magazines.

These latter data enable assessment systems to report test results in line with student socioeconomic and cultural background, and thereby make informed judgments about the quality of schools according to students' degree of social vulnerability. Of note in this approach are the experimental efforts made by some subnational systems, such as Buenos Aires Province in Argentina, to measure value added. Using performance tests administered at two different times, this model makes it possible to assess the progress made in a given period by a cohort of students from the same socioeconomic background. Unfortunately, the exercise in Argentina was discontinued, and the assessment results have not been published.

The information generated by the national and subnational assessment systems has three main audiences or targets:

- government authorities in the executive and legislative branches and the senior management of the education sector in each country;
- the general public, as well as academic and nongovernmental organizations; and
- local educational actors, primarily teachers, school principals and supervisors, students, and parents.

Throughout the 1990s, national assessment systems developed different formats for their results reports, with the

## **Results Reporting**

---

<sup>15</sup>The information collected via these questionnaires is not always useful for statistical analysis partly because the statistical methods used are insufficiently robust to allow for rigorous analysis, and partly because it is difficult to "measure" pedagogical practices and other relevant variables on a large scale.

aim of making them more useful **for** their intended audiences. There has been a general trend toward formats that increasingly seek to deliver results at the school level and that place particular emphasis on presenting data in a way that is useful for curricular and pedagogical purposes. This contrasts with the initial inclination in some countries—especially those whose assessment systems were instituted at the start of the decade—to produce general reports that featured aggregate data on overall performance percentages by subject and comparative percentages among the various student populations being tested. The utility of this approach was confined to selecting groups or sectors that could be targeted for some form of intervention or, more commonly, to mobilize public opinion on the issue of deficient educational attainment.

***Reports for Government Authorities and Senior Sectoral Management.*** All countries in Latin America produce general reports, in the form of executive summaries, on the academic performance of the students tested. The main goal of these reports is to provide information to political authorities and senior specialists, thereby facilitating education policymaking based on indicators related to educational processes and outcomes. These indicators serve to complement traditional indicators—such as wages, infrastructure, and facilities—and other system outputs—such as admission, repetition, dropout, and graduation rates.

In general, these reports to higher authorities offer a comparative description of results among the different populations tested. The academic performance data are aggregated by subject and grade at the national level, and then disaggregated by geographic or administrative regions and by school type (public or private, urban or rural, single- or multi-teacher, bilingual or monolingual, etc.). Some reports disaggregate the data by the socioeconomic or sociocultural circumstances of the students being assessed. The information is normally presented via graphs, tables, and short narrative summaries. The reports almost always include some general conclusions that compare subgroups but avoid giving specific figures, and they generally conclude with a series of recommendations

on possible courses of action to improve educational processes or enhance equity in access to knowledge.

Those conclusions and recommendations are normally based on an analysis of the in-school and out-of-school variables that are statistically correlated with academic performance. As noted earlier, however, there are constraints to the validity of the interpretations that can be made on the basis of such analyses. As Ravela (2001a) notes in his study of national results reports, the validity of interpretations of performance-related factors is undermined if they are based solely on bivariate analyses, since the latter do not take account of other factors that simultaneously affect performance and alter the impact of the other variables.

***Reports to the General Public.*** Reporting to the general public is less consistent across countries, and, in general, insufficient time and resources seem to have been invested in developing formats geared to that end. High-stakes assessment systems, which must systematically give information to direct users of the data, make results publicly available by school or by student. Such is the case of SIMCE in Chile and ICFES in Colombia. In countries where tests do not have direct repercussions for local actors, efforts to deliver results to the general public have been intermittent at best, and access to the data depends mainly on the willingness and capacity of the media to cover the issue with some regularity.

***Reports for Local Education Actors.*** Much remains to be done to improve the way in which the results reports are designed and how information is conveyed to local actors. Nonetheless, many efforts have been made in this regard, as evidenced by the variety of formats different countries have produced. Many offer very useful data, such as substantive analysis of student answers, and pedagogical and curricular recommendations for teachers and students. Some national and subnational cases are presented below by way of example.

**High-Stakes Exams.** In recent years, countries that administer tests to high school graduates or for admission to higher education have devised new conceptual models of assessment and report formats. These provide interest-

ing and useful data for teachers and students that go well beyond the mere transcription of scores and class rankings used for accreditation or student selection. In some cases, these models are more closely aligned with the curricular content of secondary education, which is in contrast to the classic assessment of cognitive skills and the establishment of achievement criteria that make it possible to assess students relative to agreed learning targets, rather than obtaining simple measures of central tendency (normative tests).

In the Dominican Republic, the institutional reports given to the schools disaggregate results by course or section and by knowledge area. They note the percentages of students who move forward a grade and who are kept behind, and the correlation between final school score (internal) and the score in the standardized test; they also provide a comparison of results with similar schools, as well as with all schools in the country, region, and district. The students receive an individual report of their score by domain or skill, as well as by performance level, in each knowledge area. They are also given a narrative analysis of their main achievements and difficulties, and a table that compares the results to both the national average and the average of schools at the same socioeconomic level or in the same geographic area. Finally, they receive a series of specific recommendations on how to improve their academic performance; this is especially useful for those students who failed the test and have to take it again.

The reports on the Learning and Aptitude Test for High School Students (PAES) in El Salvador also give results to students individually. Those reports provide every student with information on the skills tested in each subject, the extent to which those skills have been attained, the overall performance average in each area, and a conceptual explanation of the levels of attainment reached.

The results of Colombia's state examinations are given to all schools and to students individually. All high schools receive an institutional report that details the achievement level of each of their students in the various areas of the curriculum and explains toward which specific goals in each area there has been the greatest and least progress. Students

receive individual reports informing them in detail of their most significant achievements and difficulties.<sup>14</sup>

The preceding cases involve tests that have high stakes for the students but not the schools, at least formally. Information gathered for this study indicates that only in Chile are the results of the tests for admission to higher education made available on an individual, school-specific basis.<sup>15</sup> In this case—and since some parents and students can choose the public, private, or subsidized school they prefer—the institutional results can have direct consequences for secondary schools in terms of the students they are able to enroll.

**Low-Stakes Exams.** Most assessment systems that provide results with no direct consequences for institutions or individuals do so with the aim of offering local actors inputs that enable them to reflect on student performance and devise improvement strategies to address any apparent difficulties. Some countries and subnational systems have made assiduous efforts to develop formats for those purposes, including Bolivia, Uruguay, Aguascalientes, Bogotá, Minas Gerais, Paraná, and São Paulo.

In Uruguay, data disaggregated by school are given confidentially to each establishment; they present student results and those of students in schools in similar socioeconomic circumstances. Data on academic proficiency is presented based on a cut-off line (satisfactory performance) that is equivalent to correctly answering 60 percent of all test questions, and the percentage of students who reach that level. The reports on the sixth grade of primary school also offer multiyear comparisons (three-year periods) of the extent to which the student population has attained the skills being tested. The data are given as percentage points of difference between the results of the comparative tests (years 1 and 3) for the entire population being tested, and are disaggregated by socioeconomic context. Perhaps the main value of these

---

<sup>14</sup>This kind of report can be viewed as an important step toward the type of instruments and report formats advocated by the standards movement.

<sup>15</sup>Efforts were made to publish the comparative data in some countries, but such initiatives did not become permanent features of the dissemination process, largely because the education community resisted the publication of institutional rankings.

reports is that they present a selection of test items; teachers can therefore replicate them independently during the school year, following the grading guidelines provided by the reports. Schools that were not part of the national sample can thus secure a more objective measure of their students' performance level, gain access to new assessment methods, and obtain an opportunity for more systematic reflection on the curriculum and on teaching-learning processes. Additionally, they have a chance to devise initiatives and receive support to upgrade their institutional plans and programs.

Bolivia's System for Measuring and Evaluating the Quality of Education (SIMECAL), for both its census-based and sample-based tests, gives the results to schools in the form of an institutional report on their students' performance, one that includes the average institutional score, an operational description of achievement levels by area, and the percentage of students in the school at each level. This information is followed by a description of the strengths and weaknesses of the entire student population by core topics in each area of the curriculum. Schools also receive a pamphlet containing an analysis of the chosen items and methodological proposals for improving the results associated with those questions. Graduating high school students are each given an individual report on their performance in each core topic assessed by area. All schools and local technical teams receive a pamphlet with indicators of school effectiveness in line with the information gathered through the tests and the context questionnaires. The published performance-related factors refer to the characteristics of the student population and their families (previous education, socioeconomic status and attitudes, housing), teaching practices, and the school's material resources. Information is available on families' socioeconomic and cultural characteristics, but no specific strata have been defined that would allow the results to be disaggregated by these variables.

In the Mexican state of Aguascalientes, the results of the state tests are reported by performance levels in each area. The levels correspond to a four-way division on a scale of 1 to 100 and are classified as critical and insufficient (non-domain) and acceptable and desirable (domain). At the end of each academic year, the results are presented in terms



of educational gain on a scale of -4 to 4, depending on the progress and setbacks made apparent by a comparison of the initial and final tests. For instance, if a student began the year at an insufficient level and ended it at a desirable level, the gain—according to the formula used—is calculated at 3 out of 4 points. If the opposite were the case—that is, if a student began at a desirable level and his or her performance deteriorated to insufficient—the gain is expressed in negative terms, that is, as -3 points on the scale. These data are aggregated at the state level and disaggregated by educational area, school, group (course), and individual student.

In the reports the Bogotá education secretariat distributes to every school, the results for each subject are disaggregated by skills and by performance levels in each skill; they present the expected value (official expectation) and average value that the students attain on a scale of 0 to 306. The reports also analyze the error patterns for a selection of items (for all schools), and offer a series of teaching recommendations based on the data provided.

In Minas Gerais, every school receives an individual report. These provide separate graphs for all grades and curricular areas tested, and every graph shows the percentage of students in each of the three predetermined performance (proficiency) levels—critical, intermediate, and sufficient. The percentages are disaggregated at the level of the different abilities that comprise each of the skills being tested in each curricular area. The reports also include a bar chart that compares the school's average performance relative to that in the corresponding municipality, regional superintendency, and state. The reports are complemented by an explanatory guide to help teachers understand and interpret the data.

In the states of São Paulo and Paraná, each school takes responsibility for the primary processing of the data (counting responses, basic descriptive statistics, etc.) in line with the format and guidelines provided by the secretariat's assessment unit. These data are submitted to the central administration, where they are again processed and verified. They are returned to the schools in a new format and complemented by methodological guides that interpret the results. In Paraná, each school is entrusted with producing a

school bulletin that—in addition to the test results—includes a series of complementary institutional data such as pass, failure, and dropout rates, and the teachers' professional profile. These data are also processed by the assessment unit and returned to the schools in a format that facilitates comparison with other schools in the municipality and throughout the state.

**Reports for Parents.** In almost all countries of the region, this target group is the poorest served by assessment systems. Perhaps the main reason for this weakness is educational authorities' conviction that such information might prompt comparisons and interschool competition for enrollment. Another reason often given is that parents are not able to understand the information, making the effort involved in divulging the results unwarranted. Both reasons are highly debatable, of course, but for now this debate seems to have no place in the discussion of assessment and the way in which achievement information is used.

In most countries with census-based tests, schools are asked to share the results with parents, but this is simply recommended, not mandated. Chile is the only country that systematically grants families access to the results, as part of its system of free school choice. The subnational cases of Paraná and Minas Gerais are also worth noting in this context. In the former, groups of parent representatives are invited regularly to state meetings on teacher training, where they are informed of the annual test results and their implications for proposed policy initiatives. Not all families in Minas Gerais receive results reports, but reports are given to family representatives serving on the governing boards of each school. These representatives have access to the reports and can make them, or the information they contain, available to other parents at their discretion.

**Qualitative Follow-ups.** In addition to the regular reports discussed above, some countries undertake qualitative monitoring of the institutional and pedagogical processes that might explain student results. The Curriculum and Assessment Unit of Chile's Ministry of Education, in which SIMCE operates, has a special monitoring team that conducts sur-

veys and carries out observations in the schools. These activities seek to determine the scope of the new curricular frameworks and the extent to which they are being implemented, as well as to appraise the use of mass-distribution schoolbooks and teacher styles of instruction. The findings of this research are usually presented as percentage responses to survey questions. Since the sample of schools in these monitoring exercises is not statistically representative, the data are not presented as correlations among process and outcome variables, but as trends that should be studied in greater depth in the next national tests. As far as is known, this information is for the ministry's internal use in devising curricular policies and has not been shared with the members and institutions of the academic community.

Recently in Mexico, the Secretariat of Public Education's General Assessment Directorate conducted qualitative monitoring of the schools in the national sample. The results of these exercises were given to the schools in reports that describe the organizational and pedagogical characteristics of those schools whose results improved over time, as well as those whose performance deteriorated. They also provide methodological guidelines as to how such research can be reproduced in each school through self-assessment, and include a series of pedagogical and organizational recommendations to improve educational quality.

This area is one of the most problematic for the region's assessment systems. The national systems were created with the explicit goals of making the outcomes of educational management and practices in schools more transparent and of providing professionals with useful tools to improve their practices. The national education authorities have thus assumed a very grave responsibility, which they do not always fulfill in a timely and appropriate manner.

As explained earlier, test results are put to varied uses that can be classified broadly as high stakes or low stakes, depending on the consequences or direct accountability they entail for the system's local and central actors. The way in which the information is to be disseminated, and the strategies for doing so, largely depend on the expected impact

## **Dissemination and Use of Results**

of the data. Ideally, and regardless of the intended audiences and outreach strategies, the expectation is that test results will cease to be mere data and will instead become substantive information on students' educational achievement. This means that those responsible for disseminating information must also assume responsibility for interpreting it, reorganizing it, and presenting it in a clear manner that distinguishes among the various uses to which different audiences are expected to put it. In other words, the data become information when they intentionally send a message about educational performance to the actors in the system. Ideally, that message should be consistent with a conceptual framework that makes it possible to interpret results according to a series of clear criteria on what students are expected to learn and how they are to perform.

Other factors that affect the prospects of the data having an impact are the time periods within which the results are provided and the regularity with which they are delivered. That is, actors in the education community will be able to put the data to better use if the results are conveyed to them shortly after the tests, and if they receive the information regularly. These factors can help consolidate an assessment system, since they spur growing demand for timely data in those crucial periods, when educational improvement policies are being devised or when the coverage and delivery of the curriculum in the schools are being planned. Several of the region's assessment systems have administered tests and delivered the results in regular cycles, usually annually or biennially. In other cases, the tests and the delivery of the results have been intermittent, or have been regular for a time and then discontinued.

In any case, it cannot be said that if the data were made available more regularly and consistently they would have a greater impact. The impact depends on other important factors, such as the participation of local actors in the assessment process, the quality of the reports, and political will to make informed (not simply intuitive) decisions. Moreover, it is important to distinguish between regularity and frequency in testing and results delivery. For instance, annual testing and delivery do not themselves guarantee a greater impact

on the education system, since such a pace normally leaves the technical assessment staff with less time to undertake substantive analysis of the data and thereby improve the reports. Hence several of the region's assessment systems have decided to test less often but to continue to test regularly, so as to provide their audiences with better information. Peru is an example in this respect. The original plan had been to administer national tests every two years in several grades and disciplines, but the time frame was expanded to three years. In Chile, too, the regularity with which the tests are held has been changed: the tested grades and subjects alternate, so that more time and resources are available to analyze the data and improve the system in general.

***Use in Policymaking or in Designing Improvement Programs.***<sup>16</sup> Only a few countries have used test data in education policymaking; these include Chile, Mexico, and Uruguay.

In Chile, the SIMCE assessments are used to target improvement policies in three ways:

- The results are one of the main criteria used to choose schools to be included in the 900 Schools Program, which currently serves about 1,500 institutions.
- The data are used to allocate competitive funds for educational improvement projects in schools.
- The data comprise one of the main indicators for granting teaching incentives through the National Performance Assessment System for educational establishments.

In Mexico, a school improvement initiative known as the Quality Schools Program began recently among schools in marginal urban areas.<sup>17</sup> Several states have chosen the

---

<sup>16</sup>This section excludes use of assessments for accrediting or selecting students, or use by students and their families for the purpose of choosing schools.

<sup>17</sup>This program is coordinated at the federal level and involves research teams in each of the states. It consists of qualitative research in the schools on the institutional factors that help or hinder the students' performance. The program calls for the schools to make a proposal on improving results, and the initiatives are monitored during the implementation phase in the form of qualitative analysis.

schools that take part in the program on the basis of the socioeconomic context in which they operate, as well as their results on national and state tests. These same tests, together with other indicators and ad hoc assessments, are used to monitor the progress of the participating schools.

The tests in Uruguay were used recently to support the continuation and expansion of an equity program for full-time schools. Because of the testing model used in Uruguay, the data serve to show that such schools attain better results than others in similar socioeconomic circumstances.

Polymaking on the basis of test results is more extensive at the subnational level, especially in developing teacher training and school management programs. Notable in this regard are the assessment systems in Bogotá, Paraná, and São Paulo. All of these feature comprehensive and continuous training programs for teachers and principals designed to improve student results in local standardized tests.

In Bogotá, there are reports that many of the decisions on investment in school infrastructure, school meals programs, and other matters take test results into account. The data are also used to select and monitor the private schools that accommodate public enrollment through per capita subsidies.

***Local Pedagogical Use.*** Some learning assessment systems seek to give schools useful pedagogical tools to help them improve student academic performance by producing analytical reports on the results of each test and context questionnaire, and by delivering the results in workshops.

In Uruguay, workshops are held that seek to prompt debate among professionals on the results for various test items and the validity of those items. The delivery of the results is geared not simply to improving scores but to ensuring that scores reflect a new educational outlook. After the census- or sample-based tests, interpretation guides are distributed with a copy of the complete test. This allows teachers to reproduce the centrally administered test and grade them individually. Their ownership of the tests also allows them to design similar instruments independently. It is important that teachers receive the data shortly after the test, a process that usually takes one to two weeks.

To deliver the results of Colombia's state examination, training workshops on interpretation of results have been organized for 1,400 schools. According to information from 2001, there were plans to extend the program to 6,600 more schools throughout the country. The workshops provide training on how to interpret institutional reports and how to identify each school's strengths and weaknesses. They last two hours and are offered to 80 school representatives at a time. Before the adoption of the new, criterion-referenced tests, 12 workshops for 11,000 school representatives sought to help teachers understand the tests' conceptual framework and structure.

In Bogotá's subnational assessment system, the dissemination process begins before the tests are administered; measures are taken to explain the procedure to teachers and sensitize them thereto. Before the first general test, a pilot test was conducted to promote interest. Once the results of the first general tests were available, interpretation workshops and seminars were held for 6,000 participants, to which each school in the district sent its principal, two or three teachers, and its academic coordinator.<sup>18</sup> Videotapes and DVDs are used to present the results to the educational community and general public.

Bogotá's education secretariat additionally devised two intervention programs. One of these, Leveling for Excellence, provides support for management, infrastructure, and teaching resources in those schools with the poorest results. For these interventions, especially the teaching component, universities are contracted as training providers that work on site with teachers and principals. The training is conducted through periodic visits to the school over the course of a year. The most recent assessments indicate that academic performance in the participating schools rose by 35 percent between 1998 and 2000, while the average for Bogotá as a whole grew by only 11 percentage points. In the other intervention program,

---

<sup>18</sup>The scope and outcome of the seminars is assessed through surveys on the extent to which participants were satisfied with the events. According to recent surveys, beneficiaries of the course demonstrate high levels of approval, as they now understand the contents of the reports.

Action for Excellence, those public schools with the best results (upper decile) exchange information on their experiences; data on their practices are collected; and publications on those experiences are distributed to the schools.

**High-Stakes Use.** Some cases of high-stakes assessments were described earlier, mainly those that regulate admission to higher education or graduation from basic secondary education. Other assessment systems are used to establish teacher incentives, allocate extraordinary financial resources, or monitor and regulate student advancement to higher grades in basic schooling.

Both Chile and Mexico use student test results as complementary indicators for teacher incentives. In Chile, this is done through the National Performance Assessment System for subsidized schools, which gives an annual bonus to the teaching staff in schools that (in addition to other professional achievements) raise the academic performance of their students relative to previous assessments. The SIMCE results are also used to select and monitor the progress of schools that seek financing for Educational Improvement Projects, a system of competitive funds to develop and implement institutional improvement initiatives. In Mexico, teachers can voluntarily apply to the Teaching Career Program, a system of wage incentives that uses national test results as an indicator of professional performance.

São Paulo is an example of a system that uses results to advance students to higher grades. Here, census-based SARESP data are used as an indicator of whether students should advance to the next grade of basic education. Deficient performance in the SARESP tests (less than 50 percent correct answers) can be revised by the authorities if a student's performance had been very good in the rest of that grade, and, for some reason, he or she was unable to perform to his or her full potential on the day of the test. In such cases, students can take the test a second time.<sup>19</sup>

---

<sup>19</sup>As of this writing, the education secretariat is reconsidering use of the SARESP assessments for the purposes of giving wage bonuses and advancing students to the next grade.



**Political Accountability.** Official educational assessment documents often state that test results are meant to establish political and administrative accountability for the system's achievements. In practice, however, there are few cases of genuine state accountability to society and to the educational community for educational outcomes in terms of academic performance. There are some legitimate reasons not to make results widely available as a critical indicator of the success of the sector's administration. These include the fact that the results are conditioned by factors the education system cannot control—for example, the data secured at a particular time reflect the outcome of an educational process that lasts longer than a normal presidential or ministerial term. Nonetheless, the most modern assessment systems are able to measure educational progress from one year to the next, controlling for out-of-school variables. In several countries, however, results are seldom made available in formats or media that facilitate public scrutiny.

At least two cases are exceptions in this regard. One is Chile, where results are published every year by SIMCE and normally have significant political repercussions. This is evidenced mainly by academic articles and opinion pieces that use the results and the assessment databases to question the outcomes of educational reform in terms of quality and equity.

Another notable case is Bogotá, where the education secretariat publicly announced that its aim is to ensure that, by 2004, all schools in the district would score the 180-point average (on a scale of 0 to 306), and that the overall minimum performance would be 100 points in each school. No information is yet available on whether these goals have been reached, nor on their political repercussions, but the case is noteworthy because it is the only system that makes the aims of its education policy explicit in terms of educational attainment.

In recent years, another important aspect of the development of national assessment systems in Latin America has been countries' participation in international educational achievement tests.

**Participation  
in International  
Tests**

Comparative international tests are standardized instruments administered simultaneously in several countries to selected grades or age groups. They gather conceptual or procedural information on educational achievement in various disciplines as well as on the contextual factors that are assumed to influence students' academic performance. The tests normally include multiple-choice items, open-ended questions, and/or practical demonstrations (such as laboratory exercises in the natural sciences). Contextual factors considered normally include students' family characteristics and socioeconomic background, school infrastructure, teachers' academic background, the extent to which classroom teaching covers the curriculum in the disciplinary areas tested, and the attitudes and values students demonstrate in those areas (Ferrer and Arregui 2003).

The tests are designed and coordinated by international assessment agencies with some involvement from the individual countries' educational assessment units. Some of those agencies are private, independent organizations primarily focused on education, such as the International Association for the Evaluation of Educational Achievement (IEA) and the Educational Testing Service (ETS). Others are intergovernmental agencies that elicit participation from member countries' political and/or technical authorities for managing and assessing education. Examples of these organizations are the Organisation for Economic Co-operation and Development (OECD); the Latin American Laboratory for the Evaluation of Educational Quality (LLECE), which is an agency of the United Nations Educational, Scientific and Cultural Organization's Regional Office for Latin America and the Caribbean (UNESCO/OREALC); and the Southern and Eastern African Consortium for Monitoring Educational Quality (SACMEQ). The IEA and OECD tests have worldwide coverage, while those of such entities as LLECE and SACMEQ are regional. As of this writing, 10 separate international tests assess student achievement in a variety of knowledge areas.

Latin American countries that have participated, or are participating, in one or more of these international assessments are Argentina, Bolivia, Brazil, Chile, Colombia, Cuba, the Dominican Republic, Honduras, Mexico, Paraguay, Peru,

Uruguay, and Venezuela. In addition to those international assessments already mentioned, these countries have variously participated in the Adult Literacy and Lifeskills (ALL), International Adult Literacy Survey (IALS), Progress in International Reading Literacy Study (PIRLS), OECD's Programme for International Student Assessment (PISA) and the related PISA Plus, Trends in Mathematics and Science Study (TIMSS), and Trends in Mathematics and Science Study-Repeat (TIMSS-R). Their experiences have varied, with countries deriving greater or lesser benefits depending on their technical capacities when they embarked on the project as well as on the political conditions existing at the time. Those conditions—specifically, the political will to publicize results and the stability of the involved political and technical authorities—have markedly influenced the kind of impact the assessment results have on policy decisions and on educational practices in the schools. In general, the impact of international educational assessments in Latin America has been very weak; this is primarily due to a lack of appropriate strategies for communicating and using the results, and—as with domestic assessments—to a deficient definition of assessment purposes and goals.

## Conclusions

Following are some general conclusions about the most significant advances and challenges in measuring and assessing the quality of education in Latin America.

### Assessment Goals

In general, and in line with the cases reviewed in this study, there seems to be substantial consistency between decisions on the groups to be tested and the kinds of teaching and policy decisions the assessments are meant to inform. For instance, it is clear that countries using sample-based tests only seek to understand certain aspects of how the system performs—either to design some general policy measures or to promote political accountability for educational improvement (low stakes). Systems that use census-based tests have two main goals. On the one hand, they seek to foster accountability among schools, professionals, and students by using incentives/disincentives or accreditation mechanisms that directly affect local actors in the system (high stake). On the other hand, they seek to provide highly disaggregated information on student performance to facilitate the planning of particular curricular and pedagogical interventions. This parity notwithstanding, countries still need to determine whether the decisions taken thus far in this field truly reflect the pedagogical and curricular development needs of the system in all its depth and diversity. Specifically, it is worth considering whether the aim of every test is to ensure greater coverage of the curriculum or greater coverage of the student population, assuming that an increase in one kind of coverage constrains the prospect of increasing the other,<sup>20</sup> and that the two alternatives serve different purposes.

### System Organization

Countries have explored a variety of alternatives for organizing and administering their assessment units. The most common form of organization consists of units that are directly dependent on education ministries, semiautono-

---

<sup>20</sup>See chapter 1 of Ravela (2001b).

rious agencies (institutes), and universities contracted by the government. As noted earlier, no one form of organization can be said to be better than another. The choice depends on each country's political context and its technical and financial capacity for large-scale testing. Progress derives, in particular, from experimentation with alternative methods of managing assessment, which suggests that there is a marked willingness to maintain and improve evaluation systems even though the impact as of this writing has, in some countries, been relatively limited. It should be noted that systems managed by national institutions—that is, without the intervention of institutes or semiautonomous organizations—seem to be more viable and effective in subnational education systems (i.e., at the state or provincial level). Nonetheless, some national systems have managed to secure high levels of institutionalization by being administered directly by the education ministries.

Substantial efforts have been made to establish clearer and operationalized expectations of the kind of academic skills students should develop and hence on what they should be tested. The most concerted efforts have been made by the assessment teams themselves. These have, in the absence of clear curricular referents (as delineated in Ferrer et al. 1999 and Ferrer 2004), proposed more specific conceptual frameworks and achievement indicators that facilitate substantive interpretation of test results.<sup>21</sup> The most apparent weakness in this area consists of the scarce links between the teams responsible for assessment and those entrusted with curricular development in the education ministries. As other groups of professionals (such as those responsible for assessment) continue to make more sophisticated proposals on the kind of expectations to be assessed, it is likely that the curricular development teams will have stronger incentives to make their own proposals on how to update and design the curriculum.

## **Assessment Expectations**

---

<sup>21</sup>In terms of the move toward standards, it is not sufficient for assessors to define expectations for achievement. Those expectations should be known to all stakeholders, and the desired or anticipated performance levels fixed within a broader curricular policy, and not used solely to design models for assessing learning.

**Reporting**

The way in which information collected during assessment exercises is reported is not always entirely useful for the intended audiences. In particular, some countries report different groups' results without taking into account their economic and sociocultural characteristics. It is unfair to compare the academic performance of schools or localities without making it clear that in some of them the student population is highly disadvantaged, making it more difficult for them to achieve results comparable to more advantaged schools. Several countries are attempting to report data in such a way that schools can compare themselves with similar institutions, thereby securing a fairer measure of the efficiency of teaching efforts. Also, various (albeit incipient) initiatives are under way to measure these differences statistically using value-added models. Strategies need to be improved for informing system professionals and beneficiaries not only of student achievement levels but also of expected student performance; this latter information should be presented in concrete terms and clarified with examples. Although parents normally find it difficult to express their own expectations of the education system, evidence suggests that they are indeed interested in knowing more about schools' curricular expectations.

**Dissemination and Use of Results**

For both national and international tests, it is apparent that information exchange between assessment entities and system users remains weak, despite a variety of available reporting mechanisms. Better communication strategies are needed to enable information users—including policymakers, teachers and principals, parents, the media, professional associations, and representatives of the general public—to take part in devising tests, designing reports, and defining strategies for dissemination and use. Such strategies would make it possible for assessment data to be transformed into substantive information that is of real interest to different audiences; they in turn would respond better to assessment information if they were drawn into the process from the outset.

Countries with high-stakes systems have more robust dissemination strategies and make greater effort to deliver information, disaggregated by school and even by course and

student, to a larger number of actors. These strategies are normally part of the logic of high-stakes assessment, since they seek institutional (school) accountability and individual (teacher and student) accountability. Additionally, some systems provide results to these actors even when the information is to be used to encourage—not necessarily oblige, through sanctions and incentives—professional accountability in the interests of improving teaching and management processes.

A distinction must be drawn between systems that put greater emphasis on increasing political and administrative accountability for educational outcomes—and that promote accountability among schools, teachers, and students (high stakes)—and those that seek to have a local pedagogical and curricular impact. As of this writing, the former efforts have been less successful, since the general public's limited participation in assessment processes has not been conducive to any significant rise in demand for better quality. In other words, although the results are often published in the media, they have not prompted strong opinions among some key sectors of civil society, and thus the pressure these exert on policy decisions has been weaker than anticipated. However, in some countries, results have a greater impact than they did a few years ago, inciting at least discussion if not always action.

Countries that seek to promote institutional, professional, and student accountability have more specific normative frameworks and management mechanisms to ensure results are used broadly and systematically. Of course, that does not necessarily guarantee better educational quality; thus, in recent years, countries that use high-stakes tests have devised more sophisticated reporting formats that are more useful from a pedagogical and curriculum perspective for teachers and students. As noted earlier, the most progress in this regard has been made by systems geared to accrediting students when they leave high school or for admission to higher education.

It is often pointed out that high-stakes tests can distort teaching practices because teachers' efforts focus on improv-

## **Account- ability**

ing test scores to the detriment of imparting other formative educational values. This does not seem to be a problem in Latin America, at least as of this writing. Although this study did not seek to elicit the opinions of local actors on this matter, other research administered by the author in Latin American countries that use such high-stakes tests indicates that accreditation and incentives mechanisms do not necessarily mean that classroom efforts are reduced to “teaching to the test.” In fact, many schools with high scores on standardized tests have made their own education and curricula proposals that place a marked emphasis on the integral, formative role of their students (Ferrer 2004).

### **Subnational Assessments**

Some solid initiatives have been geared to developing systems that take greater account of local or regional information needs. Because of their scale and geographic restrictions, such systems have the potential to devise instruments that are more closely aligned to local curricular definitions (assuming that such exist) and to develop more effective strategies for communicating results. It would be helpful if a second phase of this study were to explore this issue in greater depth. As a start, it can be said here that such initiatives have achieved a significant degree of internal consolidation and legitimacy in the eyes of the educational community, and that they are significantly more likely to have an impact on curricular and teaching practices in the schools than are national systems.

### **Participation in International Tests**

About 20 Latin American countries have taken part in at least one comparative international assessment, and some countries have participated in several, over the past 10 years. Their experiences have varied significantly, depending on their technical capabilities and the political contexts in which the tests have been held; the relative advantages/disadvantages of participation have correspondingly differed. In several countries, the experience has provided the technical assessment teams with an excellent opportunity for learning and training, even if the data collected have had little impact. Mass dissemination of the results, mainly through the media, has



prompted some criticism because of the wide performance gaps between Latin America and more-developed countries. However, the results have not been circulated in the professional education media in such a way as to induce changes in pedagogical, curricular, or management practices, nor to spur decisions in those areas.<sup>22</sup> As with the national tests, communication between the international test administrators and the potential users of the assessment data needs to be strengthened before, during and after test administration. In the short and medium term, this is the only way to ensure that test results are more widely disseminated and have a greater impact on improving educational quality.

---

<sup>22</sup>In the past several years, the OECD's PISA has provoked widespread debate in the region, largely because of the academic content it proposes, its definitions of socially significant learning, and the various possible performance levels.



**Part II.**

---

**National/Subnational  
Snapshots**

*Note:* The data presented here on the publication and use of assessment results were provided by technical or advisory authorities in the respective country/district. These data might not be consistent with the findings presented in the descriptive paragraphs on each system, which are based on the author's research. Both sets of information are included to highlight the differences between official aims for the use of assessment results and what actually happens.

## Argentina

| Test year                                     | Grades tested | Subjects tested  | Test coverage   | Results reported to   | Results used for  |
|---|---------------|--|-----------------|---|---|
| Domestic assessments, administered by DiNIECE |               |  |                 |   |   |
| 1993–1994                                     | 7, 5M, 6M     | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> </ul>  | National sample | <ul style="list-style-type: none"> <li>• Government</li> <li>• Users</li> <li>• Public</li> </ul> | <ul style="list-style-type: none"> <li>• Teacher training</li> <li>• Management training</li> </ul> |
| 1995  | 3, 9, 5M      | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> <li>• Natural sciences</li> <li>• Social sciences</li> </ul> |                 |   |   |
|   | 6M, 7         |  |                 |   |   |
| 1996–1998                                     | 3, 9          | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> </ul>  | National census |   |   |
|   | 5M, 6M        | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> <li>• Natural sciences</li> <li>• Social sciences</li> </ul> | National sample |   |   |
|   | 6, 7          |  |                 |   |   |
| 1999  | 3, 7, 9       | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> </ul>  | National sample |   |   |
|   | 6             | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> </ul>  |                 | National census   |   |
|   | 5M, 6M        |  |                 |   |   |
| 2000  | 3             | <ul style="list-style-type: none"> <li>• Natural sciences</li> <li>• Social sciences</li> </ul>  | National sample | <ul style="list-style-type: none"> <li>• Government</li> <li>• Public</li> </ul>                  | —   |
|   | 6, 5M, 6M     | <ul style="list-style-type: none"> <li>• Natural sciences</li> <li>• Social sciences</li> </ul>  | National census |   |   |
|   | 9             |  |                 |   |   |
| 2002  | 3, 6, 9, 5M   | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> </ul>  | National sample | —   |   |

| Test year  | Grades tested       | Subjects tested   | Test coverage   | Results reported to | Results used for |
|--|---------------------|---|---|---------------------|------------------|
| Domestic assessments, administered by DiNIECE  |                     |   |   |                     |                  |
| 2003   | 3, 6, 9, 12         | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> </ul> | National sample   |                     |                  |
| 2005   | Pilot tests<br>3, 6 | Social sciences   | —   | —                   | —                |
|  | 9, 12               | Natural sciences  |   |                     |                  |
| International assessments  |                     |   |   |                     |                  |
| <ul style="list-style-type: none"> <li>• LLECE (1997)</li> <li>• TIMSS-R (2000, 2003)</li> <li>• PISA Plus (2001)</li> </ul> |                     |   | <ul style="list-style-type: none"> <li>• PIRLS (2001)</li> <li>• IEA-Civic Education (2001)</li> <li>• PISA (2003)</li> </ul> |                     |                  |

### **Institutional Framework**

Argentina's Ministry of Education, Science, and Technology has been assessing educational quality since 1993. Until 2001, its assessment agency was the National System for the Assessment of Educational Quality, which relied on input from the Secretariat for Educational Planning and Management and the Subsecretariat for the Assessment of Educational Quality. In 2000, the government established the Institute for Educational Quality, a semiautonomous agency of the Ministry of Education. The institute was dissolved shortly thereafter and has been replaced by the National Directorate for Information and Assessment of Educational Quality (DiNIECE), a line agency of the Ministry of Education.

### **Curriculum and Standards**

In 1994, Argentina approved the Common Basic Contents (CBC), as part of a reform and decentralization process in which the management of all national schools was transferred to the provinces (federal states). Each province, and the federal capital, is responsible for drawing up its own curriculum, which should be consistent with the CBC. Core learning goals are now being prepared for all levels of schooling. These goals, which are based on the CBC and approved by federal consensus, prioritize certain educational content in mathematics, language, and the social and natural sciences.

Every year, the performance of representative samples of students is assessed, and all students are tested in the final year of high school. The instruments used are norm-referenced, based on multiple-choice questions; they are complemented by questionnaires to students, principals, and teachers aimed at identifying performance-related factors. As of this writing, a new criterion-referenced model is being designed and will be applied for the first time in 2006.

### **Types of Tests**

Assessment results are disaggregated by subject, gender, province, geographic area (urban or rural), and school management (public or private). Students are classified according to certain indicators of socioeconomic vulnerability, and schools according to their material and teaching resources. For the census-based assessments, reports are disaggregated by school. The results are reported as a percentage of correct answers for all students and for the different strata assessed.

### **Types of Reports**

Argentina's assessments are low stakes, since the results do not have direct consequences for the actors in the system. (In 2000, however, school-specific results at the secondary level were made public.) Until 2000, notebooks were published with methodological recommendations for teachers; these discussed some of the test items and explained the difficulties students had encountered in the tests. These notebooks are no longer produced, however, and assessment results do not seem to have found a place in either public debate or the agenda of the education system. Only international test results seem to have some impact on the media and public opinion.

### **Use and Dissemination of Results**

The province of Buenos Aires, with technical assistance from UNESCO/OREALC, has developed an assessment system based on the "value-added" model which makes it possible to evaluate students' academic progress while taking account of the socioeconomic characteristics of each school's student population. The federal capital also has its own assessment system, but the results are not made public. Instead,

### **Subnational Systems**

they are given to the schools for research purposes, and for pedagogical and institutional improvements. Independent assessments have also been carried out in the provinces of Catamarca, Chaco, Córdoba, Entre Ríos, La Pampa, La Rioja, Mendoza, Misiones, Neuquén, Río Negro, Salta, San Juan, Santa Cruz, Santa Fe, Tierra del Fuego, and Tucumán. Coverage and methodology vary by province, as does the degree of support the provinces have sought from the Ministry of Education to develop their own instruments. Almost all of Argentina's provinces have devised some type of strategy to disseminate the results of the national assessment.

**International Tests**

Argentina has taken part in LLECE (1997), TIMSS-R (2000, 2003), PISA Plus (2001), PIRLS (2001), IEA-Civic Education Study (2001), and PISA (2003).

**Progress Observed**

Progress in Argentina includes the political legitimacy of the national assessment system, the continuity of the assessments' annual administration, and a growing national appreciation of the need to improve the utility of the assessments for teaching purposes.

**Difficulties Observed**

The main difficulties currently facing Argentina's educational assessment system include a lack of communication between the assessment unit and the offices responsible for curricular development and teacher training, and resistance to quantitative assessment on the part of some academics and technical specialists within the ministry. The assessment unit should undertake a more complex analysis of the data gathered and report the results in such a way as to distinguish among schools from different socioeconomic strata. There is no conceptual framework for the interpretation of results or performance-related factors. Efforts have been made, however, to present the data in a more substantive way so as to show the specific abilities demonstrated by all students in accordance with the level of achievement revealed by the assessments.



The primary aims of Argentina's Educational Quality Assessment Plan 2003–2007, which was recently approved by the Federal Education Council, are to develop new quantitative and qualitative technologies, undertake further analysis and interpretation of national assessment results, offer support to strengthen provincial assessment systems, and forge stronger links with the ministry's other programs and directorates to develop more effective plans for improving educational quality. Assessment information needs to be better used so as to devise policies that enhance equity in access to knowledge; also, different kinds of reports need to be developed for different audiences. The country will continue to take part in international tests.

## **Outlook**

## Bolivia

| Test year  | Grades tested  | Subjects tested   | Test coverage  | Results reported to   | Results used for |
|--|----------------|---|--|---|------------------|
| Domestic assessments, administered by SIMECAL  |                |   |  |   |                  |
| 1996–2000  | 1, 3, 6, 8, 4M | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> </ul> | <ul style="list-style-type: none"> <li>• National sample</li> <li>• National census</li> <li>• Regional census (La Paz)</li> </ul> | <ul style="list-style-type: none"> <li>• Government</li> <li>• Users</li> <li>• Public</li> </ul> | —                |
| International assessments  |                |   |  |   |                  |
| <ul style="list-style-type: none"> <li>• LLECE (1997)</li> <li>• Adult Literacy and Lifeskills (2001)</li> </ul> |                |   |  |   |                  |

### Institutional Framework

Bolivia's SIMECAL—System for Measuring and Evaluating the Quality of Education—was originally established to assess learning, but it has slowly assumed additional assessment responsibilities, including developing an academic aptitude test for graduating high school students, assessing substitute teachers, awarding teaching diplomas, managing pedagogical advisors and teachers (the bonus and promotion system), preselecting candidates to be district principals and school principals, and admitting individuals to teachers' colleges. By law, SIMECAL is an agency of the Bolivian Ministry of Education. In practice, however, it is located within the General Directorate of Sectoral Strategy. Consequently, the data it gathers in its periodic educational assessments is slow to reach the political authorities with the greatest decisionmaking power. This situation remained unresolved as of 2002.

### Curriculum and Standards

Between 1997 and 1998, Bolivia's education reform instituted a new national curricular framework for primary education that focused on skills, performance indicators, and assessment strategies by discipline. The old frameworks introduced in the 1970s remained in effect for secondary education. Arranged and disseminated in modules, the new framework does not aim for an exhaustive coverage of the curriculum but rather proposes activities that respond to a new way of thinking about the curriculum and about teach-

ing. The new curricular proposals, and many of the reform activities, have reinvigorated education in indigenous languages and cultures. Officially, the SIMECAL tests are referenced to the national curricular frameworks. However, the technical teams responsible for curricular development have expressed some criticism of the relatively isolated manner in which SIMECAL drew up the specifications tables and the assessment instruments.

Originally, the tests devised by SIMECAL were strictly normative and referred to the old national curriculum. Criterion-referenced tests were later developed on the basis of the new curricular skills and the three levels of achievement (A, B, and C) established by the assessment unit itself. The tests are made up of multiple-choice questions. They are written in the three primary indigenous languages and are complemented by questionnaires for teachers, principals, parents, and students for the subsequent analysis of performance-related factors.

SIMECAL's results reports indicate the percentages of students who reach each performance level in the areas being assessed. The data are mainly disaggregated by geographical department, indigenous language, school management (public or private), and student gender. For both census- and sample-based tests, schools receive an institutional report on their students' performance that includes the average institutional score, an operational description of achievement levels by subject, and the percentage of students in the school at each performance level. This information is followed by a standardized description of the strengths and weaknesses of the school's entire student population by core topics in each area of the curriculum. The schools also receive a pamphlet that analyzes selected questions and makes methodological proposals for improving results as they relate to those questions. Graduating high school students each receive an individual report on their performance in each core topic. All schools and local technical teams receive a pamphlet containing indicators of school effectiveness based on the infor-

## **Types of Tests**

## **Types of Reports**

mation gathered through the tests and questionnaires. The performance-related factors that are published refer to the characteristics of the students and their families (previous education, socioeconomic status, attitudes, housing), teaching practices, and the school's material resources. Information is available on families' socioeconomic and cultural characteristics, but no specific strata have been defined that would allow results to be disaggregated by these variables.

### **Use and Dissemination of Results**

The assessment data are officially presented through various channels: a public presentation by the Ministry of Education, a presentation to the various teams responsible for educational and curricular reform, the distribution of leaflets and pamphlets to the schools, workshops in capital cities involving representatives of civil society and local technical teams, and the publication of overall results in the media. Secondary studies are also carried out, linking the performance data to child nutrition variables. It is unclear what actual use the schools make of the results. The information has not been used, as had been expected, in education policymaking. Assessment results have had only a limited impact on the policy agenda due to communication problems in the ministry and, probably, the lack of agreement within the education sector regarding the validity and importance of the achievements being tested. Numerous experts in the assessment unit, as well as within international cooperation agencies, agree that the SIMECAL data should be used as one of the main criteria in deciding on the best sectoral (not only curricular and pedagogical) and intersectoral policies.

### **Subnational Systems**

Bolivia has no subnational assessment systems.

### **International Tests**

Bolivia has taken part in LLECE (1997) and ALL (2001).

### **Progress Observed**

The assessment programs have continuity, and the technical staff have been relatively stable. Gradual efforts have been made to improve the assessment instruments, devise criterion-referenced tests, and produce quality reports which,

potentially, make it possible to put the data to good use for pedagogical and curricular purposes.

The main difficulties of the Bolivian assessment system include a lack of sectoral policies that stress monitoring of educational quality and that make systematic use of SIMECAL information to that end. There is a firmly held belief that SIMECAL data could be used as a basis in intersectoral policymaking, but poor communication and inadequate links among government agencies prevent this from happening. Relatedly, there is inadequate monitoring of how schools use SIMECAL data.

### **Difficulties Observed**

The SIMECAL authorities believe that the system has taken on more assessment responsibilities than its operational capacity allows. Efforts need to be focused on assessing learning in the schools and devoting existing resources to ensure that data are distributed more extensively and that key audiences are educated on the use of the data in education policymaking.

### **Outlook**

## Brazil

| Test year  | Grades tested          | Subjects tested  | Test coverage   | Results reported to   | Results used for   |
|--|------------------------|--|---|---|--|
| Domestic assessments: SAEB, administered by INEP/MEC   |                        |  |   |   |  |
| 1990–1993  | 1, 3, 5, 7             | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> <li>• Natural sciences</li> </ul>  | National sample   | <ul style="list-style-type: none"> <li>• Government</li> <li>• Public</li> </ul>                  | <ul style="list-style-type: none"> <li>• Teacher training</li> <li>• Curricular development</li> </ul> |
| 1995   | 4, 8, 11               | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> </ul>  |   |   |  |
| 1997   |                        | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> <li>• Natural sciences</li> </ul>  |   |   |  |
| 1999   |                        | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> <li>• Natural sciences</li> <li>• Social sciences</li> </ul>   |   |   |  |
| Domestic assessments: ENEM, administered by INEP/MEC   |                        |  |   |   |  |
| 2001–2005  | 4, 8, 11               | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> </ul>  | National sample   | <ul style="list-style-type: none"> <li>• Government</li> <li>• Users</li> <li>• Public</li> </ul> | Teacher training   |
| 1998–2005  | End of basic education | <ul style="list-style-type: none"> <li>• Command of languages</li> <li>• Understanding of phenomena</li> <li>• Problem solving</li> <li>• Constructing arguments</li> <li>• Proposing socially responsible real-world interventions</li> </ul> | Voluntary   | <ul style="list-style-type: none"> <li>• Government</li> <li>• Users</li> </ul>                   | Selection for higher education   |
| International assessments  |                        |  |   |   |  |
| <ul style="list-style-type: none"> <li>• ETS-Mathematics (1991)</li> <li>• LLECE (1997)</li> </ul> |                        |  | <ul style="list-style-type: none"> <li>• PISA (2000, 2003)</li> <li>• TIMSS 2003</li> </ul> |   |  |

Brazil's national assessment agency is the Basic Education Assessment System (SAEB), which is managed by the National Institute of Educational Studies and Research. The latter, INEP, is dependent on the Ministry of Education (MEC) but has some degree of administrative autonomy. INEP has an assessment subsystem, the National High School Examination (ENEM), which has technical links to SAEB but whose tests are managed and administered independently.

## **Institutional Framework**

SAEB's first tests were established before the national curriculum standards were approved in 1997, when each state had its own curricular framework. It was thus necessary to devise, and to validate with the states, a reference matrix which contained 60 percent of the content common to all states' curricula. The matrix was later updated to include the content areas of the national curricular parameters; this meant that it was no longer necessary to validate the matrix with each state's education secretariat. At present, assessment content is defined with the help of specialists hired expressly for that purpose, although this does not amount to a joint endeavor with the stable technical teams responsible for national management of the curriculum. Teachers are consulted when the tests are drawn up, but their participation is somewhat limited. The reference matrix seeks to strike a balance between the prescribed curriculum and what the schools actually use, but the way in which the tests and questionnaires are designed precludes *ex post* analysis of the extent to which performance is attributable to the effective implementation of the national curriculum (i.e., learning opportunity).

## **Curriculum and Standards**

Brazil's testing instruments are norm-referenced and developed in line with the psychometric model of classical theory and item response theory. The reference matrix is based on descriptors involving two levels of specification: skills and abilities. The same assessment topics are used for all grades being assessed, but the priorities and levels of complexity are adjusted for the higher grades. The model is strongly cognitive, inasmuch as it seeks to determine at what point in the

## **Types of Tests**

learning process there is a break or delay in the acquisition of more complex knowledge. To that end, after the tests are administered, qualitative follow-up studies are carried out in the classroom. The tests are complemented by questionnaires that provide in-school and out-of-school data for later analysis of performance-related factors. A sociodemographic outline of the schools and their settings is created from these data, as well as a professional profile of the principals and teachers in charge of the groups being tested. The questionnaires are administered to students, teachers, and principals. The information is statistically cross referenced with data from the national education census and the census prepared by the Brazilian Institute of Geography and Statistics. The ENEM is taken by students on exiting high school. It is voluntary, and the students must register for it beforehand. The test's reference matrix was drawn up in an ad hoc manner, but it draws on SAEB's item bank and thus refers to the content of the national curricular parameters. The test collects data on indicators that can be used in selecting individuals for entry in the labor market and perhaps for admission to higher education (universities determine whether they wish to rely on these indicators).

### **Types of Reports**

The results reports mainly provide information on the percentage of students in each grade who reach different performance levels on a predetermined scale for the various areas of the curriculum. They also describe the kinds of abilities students have attained at each of the performance levels. The data are in aggregate form for the national level and are disaggregated by region, state, and municipality. The ENEM provides a student-specific results report, and anyone who desires access to this report must ask the student's permission. Schools may also request a school-level report.

### **Use and Dissemination of Results**

When the SAEB instruments were redesigned in 2001, the strategy for distributing the results and the way in which the information is used were also reconsidered. As of this writing, primary emphasis has been placed on raising awareness among key stakeholders to ensure that the information has



greater impact. Informational visits and conferences have been held in all states, involving the participation of state-level and local technical teams, teachers, principals, university personnel, and the media. At the teachers' request, these conferences stressed the technical features of the tests and the reference matrix. Assessment results are available in the public domain but are chiefly intended to be used for policymaking by the MEC. Because the tests are sample-based, results are not given to individual schools. Nonetheless, INEP has prepared specific reports for pedagogical purposes that show average student performance on each item and offer a set of possible explanations as to why students answered incorrectly. Qualitative research has sought to understand some of the factors related to teaching and social practices in the classroom that explain the results at the national level. The tests are not high stakes for the schools or students, and seem to have had no particular impact on political accountability for educational quality. The databanks and item banks are available for non-INEP research but can only be accessed following prior agreement and after legal arrangements have been made. The ENEM results are used to select those to be enrolled in the University for All Program, and for granting full or partial scholarships to students from poorer backgrounds. Hundreds of higher education institutions use ENEM results in selecting students for admission.

See the individual sections on the states of Minas Gerais, Paraná, and São Paulo, beginning on p. 132.

Brazil has taken part in ETS-Mathematics (1991), LLECE (1997), PISA (2000, 2003) and TIMSS (2003).

Educational assessment has been substantially institutionalized in Brazil, as evidenced by the stability of the teams and the continuity of the assessments. Externally, SAEB has managed to consolidate itself without triggering opposition from the trade unions or others in the education community, in large part because of its efforts to keep local actors properly informed of assessment goals. Additionally, Brazil has moved

## **Subnational Systems**

## **International Tests**

## **Progress Observed**

to an assessment model that focuses more intently on the pedagogical and curricular uses of the data produced by the tests.

**Difficulties  
Observed**

Despite the interest in maximizing the pedagogical use of test data and the efforts made in that regard, it will probably take a great deal of time (and much more dissemination and training) before the information has a real impact in the schools. No information is available regarding use of the data in education policymaking at the national level.

**Outlook**

Tailoring reports to different audiences will enable assessment results to have a greater impact on the various stakeholders involved in education.

## Chile

| Test year  | Grades tested | Subjects tested   | Test coverage  | Results reported to   | Results used for  |  |   |   |
|--|---------------|---|--|---|---|--|---|---|
| <b>Domestic assessments: PER</b>                   |               |   |  |   |   |  |   |   |
| 1982–1984  | 4, 8          | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> <li>• Natural sciences</li> <li>• Social sciences</li> </ul>                              | <ul style="list-style-type: none"> <li>• National census</li> <li>• National sample</li> </ul>     | <ul style="list-style-type: none"> <li>• Government</li> <li>• Users</li> <li>• Public</li> </ul> | —   |  |   |   |
| <b>Domestic assessments, administered by SIMCE</b> |               |   |  |   |   |  |   |   |
| 1988   | 4             | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> <li>• Natural sciences</li> <li>• Social sciences</li> <li>• Student attitudes</li> </ul> | <ul style="list-style-type: none"> <li>• National census</li> <li>• Experimental sample</li> </ul> | <ul style="list-style-type: none"> <li>• Government</li> <li>• Users</li> <li>• Public</li> </ul> | <ul style="list-style-type: none"> <li>• Teacher training</li> <li>• Curricular development</li> <li>• Targeting support to students and schools</li> </ul> |  |   |   |
| 1989   | 8             |   |  |   |   |  |   |   |
| 1990   | 4             |   |  |   |   |  |   |   |
| 1991   | 8             |   |  |   |   |  |   |   |
| 1992   | 4             |   |  |   |   |  |   |   |
| 1993   | 2M            |   |  |   |   | Student attitudes  |   |   |
| 1994   | 4, 8          | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> <li>• Natural sciences</li> <li>• Social sciences</li> <li>• Student attitudes</li> </ul> | <ul style="list-style-type: none"> <li>• National census</li> <li>• Experimental sample</li> </ul> | <ul style="list-style-type: none"> <li>• Government</li> <li>• Users</li> <li>• Public</li> </ul> | <ul style="list-style-type: none"> <li>• Teacher training</li> <li>• Curricular development</li> <li>• Targeting support to students and schools</li> </ul> |  |   |   |
| 1995   | 2M            | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> <li>• Student attitudes</li> </ul>  |  |   |   |  |   |   |
|  | 8             | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> <li>• Natural sciences</li> <li>• Social sciences</li> </ul>                              |  |   |   |  |   |   |
| 1996   | 4             | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> </ul>   |  |   |   | <ul style="list-style-type: none"> <li>• National census</li> <li>• Experimental sample</li> </ul> | <ul style="list-style-type: none"> <li>• Government</li> <li>• Users</li> <li>• Public</li> </ul> | <ul style="list-style-type: none"> <li>• Teacher training</li> <li>• Curricular development</li> <li>• Targeting support to students and schools</li> </ul> |
| 1997   | 8             | <ul style="list-style-type: none"> <li>• Natural sciences</li> <li>• Social sciences</li> <li>• Student attitudes</li> </ul>  |  |   |   |  |   |   |

| Test year  | Grades tested | Subjects tested  | Test coverage  | Results reported to   | Results used for  |
|--|---------------|--|--|---|---|
| <b>Domestic assessments, administered by SIMCE</b>   |               |  |  |   |   |
| 1998   | 2M            | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> </ul>  | <ul style="list-style-type: none"> <li>• National sample</li> <li>• Experimental sample</li> </ul>                                 | <ul style="list-style-type: none"> <li>• Government</li> <li>• Users</li> <li>• Public</li> </ul> | <ul style="list-style-type: none"> <li>• Teacher training</li> <li>• Curricular development</li> <li>• Targeting support to students and schools</li> </ul>                               |
| 1999   | 4             | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> </ul>  | <ul style="list-style-type: none"> <li>• National census</li> <li>• Experimental sample</li> </ul>                                 |   |   |
| 2000   | 8             | <ul style="list-style-type: none"> <li>• Natural sciences</li> <li>• Social sciences</li> </ul>  |  |   |   |
| 2001   | 2M            | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> </ul>  |  |   |   |
| 2002   | 4             | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> <li>• Natural sciences</li> <li>• Social sciences</li> </ul> |  |   |   |
| 2003   | 2M            | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> </ul>  | <ul style="list-style-type: none"> <li>• National census</li> </ul>  |   | <ul style="list-style-type: none"> <li>• Teacher training</li> <li>• Curricular development</li> <li>• Targeting support to students and schools</li> <li>• Teacher incentives</li> </ul> |
| 2004   | 8             | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> <li>• Natural sciences</li> <li>• Social sciences</li> </ul> |  |   |   |
| <b>International assessments</b>   |               |  |  |   |   |
| <ul style="list-style-type: none"> <li>• Six Subject Study (1970–71)</li> <li>• LLECE (1997)</li> <li>• TIMSS-R (1998)</li> <li>• IALS (1998)</li> </ul> |               |  | <ul style="list-style-type: none"> <li>• IEA-Civic Education (2000)</li> <li>• PISA Plus (2001)</li> <li>• TIMSS (2003)</li> </ul> |   |   |

SIMCE—the National System for the Assessment of Educational Quality—is administered by the Ministry of Education’s Curriculum and Assessment Unit. This system, originally created with another name outside the ministry and under the technical management of the Catholic University of Chile, has administered the country’s academic achievement tests since 1981. It began to operate within the Ministry of Education in 1988. It is not autonomous, but its technical legitimacy gives it a significant degree of leadership and decisionmaking capacity in national assessment policy.

## **Institutional Framework**

In 1996 and 1998, in one of the final phases of the broad educational reform begun at the start of the last decade, Chile’s Basic Objectives and Minimum Obligatory Contents for primary and secondary education were approved by law. These comprise the country’s official curricular framework. SIMCE’s assessment instruments gradually came to include the content of these curricular referents as the new programs were implemented in the schools. The 1990 Constitutional Law on Education stipulates that national curricular standards be established—a mandate that is now being fulfilled with the preparation of content and performance standards for all school grades in language, mathematics, history, the social sciences, the natural sciences (biology, chemistry, and physics), and English as a second language.

## **Curriculum and Standards**

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 surveying and observing teachers and principals. This monitoring facilitates more in-depth research on the impact different sources of curricular

## **Types of Tests**

design used in the schools (curricular frameworks, syllabuses, textbooks, etc.) have on students' academic attainment.

### **Types of Reports**

General reports present assessment results in aggregate form at the national level, as well as disaggregated by geographic region, type of school (urban or rural), and type of school management (municipal, private, or subsidized). Different report formats have been used in recent years, depending on the information the authors seek to impart to different audiences. Some reports present the results as averages of achievement and standard deviation for the different subjects and grades under assessment. Others provide percentages of correct answers for each subject or, at a greater level of detail, by objectives within each subject. Several reports stress the differences in results among types of school or trends in results over time. The comparability of results across years has been subject to some technical challenges, and, as a result, a standardizing methodology has been adopted that allows more reliable inferences to be drawn about the educational progress made throughout the past decade. SIMCE's technical team produces an interpretative analysis of the results and their implications to illustrate how far-reaching the implementation of the new curricular frameworks has been. All the schools that have been assessed receive an institutional report summarizing the academic achievement of the groups (courses) involved in the assessment of each subject.

### **Use and Dissemination of Results**

Among Latin American countries, Chile is probably the one in which the results of standardized tests have the greatest domestic impact. The government's efforts to promote and maintain high levels of investment in education reform have demanded that the programs implemented be seen as successful. Many of the official results reports, as well as independent research using SIMCE data, emphasize variations in educational attainment throughout the 1990s and the gaps between schools of different types (public, private) and from different socioeconomic strata. The reports are often counterbalanced, or challenged, by independent research by the

country's academic community, which questions whether the reforms and financing mechanisms have improved the quality of education. At least three strongly consolidated strategies are apparent with regard to the use of the SIMCE assessments to target improvement policies: (1) using the results as a key criterion for choosing schools to be included in the 900 Schools Program, which currently serves about 1,500 institutions; (2) using the data to allocate competitive funds for educational improvement projects; and (3) using the data as one of the main indicators to provide teaching incentives through the National Performance Assessment System for educational establishments. For pedagogical and curricular purposes, the assessments are used mainly to provide teachers with information on new and more sophisticated assessment techniques. Chile's curriculum and assessment unit also undertakes qualitative follow-up on the implementation of the curriculum in the schools. This follow-up makes it possible to observe the extent to which the new curricular frameworks are formally and effectively adopted, while the results (of both the follow-up and of the national and international tests) facilitate identification of those aspects of the curriculum that require greater impetus to be implemented and achieved.

There are no subnational assessment systems.

Chile has taken part in the Six Subject Study (1970–71), LLECE (1997), TIMSS-R (1998), IALS (1998), IEA-Civic Education Study (2000), PISA Plus (2001), and TIMSS (2003).

Originally, Chile's assessment system was a means of informing the public in choosing schools, but there has been a significant movement toward using assessment data to devise educational improvement policies. The school-specific results are widely distributed, but most of the emphasis is placed on substantive analysis of the data and their implications for improving the quality of education. The continuation of the assessment exercises over the years, the consolidation and

### **Subnational Systems**

### **International Tests**

### **Progress Observed**

stability of the technical teams, and the system's external institutionalization are also worth noting in this regard.

**Difficulties  
Observed**

Despite SIMCE's analytical reports and pedagogical recommendations, it is still difficult to affect curricular development and the teaching/learning process in schools.

**Outlook**

The current initiative to develop standards will necessitate redefining or adjusting the assessment instruments.



## Colombia

| Test year  | Grades tested | Subjects tested  | Test coverage  | Results reported to   | Results used for   |
|--|---------------|--|--|---|--|
| <b>Domestic assessments, administered by ICFES/Ministry of Education</b>                 |               |  |  |   |  |
| 1991–1994  | 3, 5, 7, 9    | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> </ul>  | <ul style="list-style-type: none"> <li>• National sample</li> <li>• Regional sample</li> </ul>         | <ul style="list-style-type: none"> <li>• Government</li> <li>• Users</li> <li>• Public</li> </ul> | —  |
| <b>Domestic assessments: SABER tests</b>   |               |  |  |   |  |
| 1997–1998  | 3, 5, 7, 9    | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> </ul>  | National sample  | <ul style="list-style-type: none"> <li>• Government</li> <li>• Users</li> <li>• Public</li> </ul> | —  |
| 1998–1999  | 7, 9          | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> </ul>  |  |   |  |
| 2002–2003  | 5, 9*         | <ul style="list-style-type: none"> <li>• Natural sciences</li> </ul>   | National census  | —   |  |
| <b>Domestic assessments: ICFES state examinations</b>                                    |               |  |  |   |  |
| 1980–2005  | 11            | <ul style="list-style-type: none"> <li>• Foreign language</li> <li>• Language</li> <li>• Mathematics</li> <li>• Natural sciences</li> <li>• Social sciences</li> </ul> | National census  | <ul style="list-style-type: none"> <li>• Government</li> <li>• Users</li> </ul>                   | <ul style="list-style-type: none"> <li>• Curricular development</li> <li>• Selection for higher education</li> </ul> |
| <b>International assessments</b>   |               |  |  |   |  |
| <ul style="list-style-type: none"> <li>• TIMSS (1995)</li> <li>• LLECE (1997)</li> </ul> |               |  | <ul style="list-style-type: none"> <li>• IEA-Civic Education (2000)</li> <li>• PIRLS (2001)</li> </ul> |   |  |

\*In some areas, grades 3 and 7 were also tested.

All educational assessment and measuring exercises are conducted through ICFES—the Colombian Institute for the Development of Higher Education—a semiautonomous agency of the Ministry of Education that began to administer national examinations in 1968. ICFES is politically dependent on the ministry but enjoys a high degree of administrative independence. As a result, it has a substantial level of technical stability and has experienced considerable institutional development over the past three decades. Within ICFES, the

### Institutional Framework

National Testing Service is the technical unit responsible for designing, conducting, and processing the national assessment exercises. The main exercise, which has been most stable over time and which is regulated by law, consists of the state examinations that govern students' admission to higher education. These tests are administered independently by ICFES and have no technical or political link to the ministry's administration of basic education (i.e., primary and secondary school). They do, however, provide information that is useful for curricular purposes and for high schools throughout the country. The ministry's Quality Directorate and ICFES have jointly developed a national educational assessment system for primary education known as the National System for Evaluating Educational Quality (SABER). SABER was conceived as a longitudinal program to monitor educational quality and was to be implemented between 1997 and 2005. Political changes and a subsequent major restructuring of the Ministry of Education, with attendant shifts in priorities and policy approaches, precluded completion of the SABER project, and the annual tests have been discontinued.

### **Curriculum and Standards**

In Colombia, three official curricular references have emerged over the past decade: curricular guidelines, indicators of general curricular achievement, and new national standards approved for three subjects in 2002. Until 1999, the SABER assessments took the first two of these as their curricular references. The state examinations were originally based on a set of universal cognitive skills, but they have recently been reformulated on the basis of criteria that take the national curricular frameworks into account. From a technical standpoint, the reference matrix for these examinations could be regarded as a proposal for standards, since it offers a clear and exhaustive explanation of what kinds and levels of academic performance are expected of students, while simultaneously serving as a useful source of information for secondary schools, students, and families.

### **Types of Tests**

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).

SABER's reports present educational results aggregated at the national level and disaggregated by department, type of school (urban or rural), and type of school management (private or public). Performance data for each subgroup of students are given as percentages of correct answers for each predefined performance level (B, C, and D—from the most basic to the most complex). The main reports for state examinations are those sent to the schools: All high schools receive an institutional report which exhaustively details the achievement level of each of their students in the different areas of the curriculum, and specifies the objectives in each area toward which the greatest and least progress has been made. Students receive individual reports informing them in detail of their most significant achievements and difficulties. Reports for mass distribution give the percentages of students who reached each of the performance levels in the different areas being assessed. They also provide data disaggregated by geographic department, comparing the performance of each to the national average.

The reports on SABER test results were distributed to the political and technical leaders in the education sector and to various local actors via pamphlets, forums, interpretative workshops, and other public events. The contextual data

## **Types of Reports**

## **Use and Dissemination of Results**

provided by the questionnaires have been analyzed mainly by external organizations and independent researchers. The results of the state examinations are sent primarily to high schools and their students. The extensive disaggregation of the data, particularly with regard to the specific knowledge that students display, makes the reports a key tool in improving curricular design and implementation. For the students, the data are precise indicators of the kind of knowledge they have acquired and of the areas in which they should work further if they are to improve their results and enhance their prospects of being admitted to higher education. There is no indication that the data have affected policy decisions on basic education, but they are commonly used as reference material by researchers from ICFES and other academic institutions.

**Subnational Systems**

The most developed subnational assessment system is that designed and implemented by the Education Secretariat of Bogotá (see p. 144).

**International Tests**

Colombia has taken part in TIMSS (1995), LLECE (1997), IEA-Civic Education Study (2000), and PIRLS (2001).

**Progress Observed**

ICFES's long experience with assessment has enabled it to develop substantial technical and operational capacities at the national level. Those capacities, however, are not always transferred to the Ministry of Education, and the ministry does not use them to inform more effective decisionmaking on education policy. Of note are the efforts made to develop psychometric models and criterion-referenced instruments, which ensure closer curricular links between high school education and university entry requirements. The internal and external institutionalization of ICFES is substantial; this is because it operates in accordance with norms and because of the technical teams' willingness to update the instruments and make them into more useful management tools.

The value of the data provided by the SABER tests and state examinations is officially acknowledged, but the results have had very little impact on education policymaking. The main reason for this is the high degree of turnover among policymakers in the education sector—a circumstance that has made it difficult to take long-term approaches and to use the assessment data to sustain those approaches.

### **Difficulties Observed**

Decisionmakers within the ministry have explicitly indicated their willingness to continue extending and institutionalizing the standards drawn up in 2002, and to strengthen measurement and assessment mechanisms in line with those standards.

### **Outlook**

## Costa Rica

| Test year   | Grades tested       | Subjects tested  | Test coverage   | Results reported to   | Results used for |
|---|---------------------|--|-----------------|---|------------------|
| Domestic assessments: Knowledge tests, administered by IIMEC and Department of National Tests |                     |  |                 |   |                  |
| 1986  | 3, 6, 9, 11/12      | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> </ul>  | National census | <ul style="list-style-type: none"> <li>• Government</li> <li>• Public</li> </ul>                  | —                |
| 1987  | 6, 9, 11/12         |  |                 |   |                  |
| 1989–1990   | 3, 6, 9             | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> <li>• Natural sciences</li> <li>• Social sciences</li> </ul>                             | National sample | <ul style="list-style-type: none"> <li>• Government</li> <li>• Users</li> <li>• Public</li> </ul> |                  |
| 1996  |                     |  |                 | —   |                  |
| 1997  | 3, 6                |  |                 | —   |                  |
| Domestic assessments: National high school exit exam, administered by IIMEC                   |                     |  |                 |   |                  |
| 1988–2003*  | High school seniors | <ul style="list-style-type: none"> <li>• Foreign language</li> <li>• Language</li> <li>• Mathematics</li> <li>• Natural sciences</li> <li>• Social sciences</li> </ul> | National census | <ul style="list-style-type: none"> <li>• Government</li> <li>• Users</li> <li>• Public</li> </ul> | —                |
| Domestic assessments: Primary school test   |                     |  |                 |   |                  |
| 1994–1996   | 1                   | —  | —               | —   | —                |
| Domestic assessments: Cognitive skills  |                     |  |                 |   |                  |
| 1996  | 1                   | —  | —               | —   | —                |
| International assessments   |                     |  |                 |   |                  |
| LLECE (1997); assessment not completed  |                     |  |                 |   |                  |

\*Assessment was conducted in earlier decades as well.

Standardized assessment activities in Costa Rica began in 1986 and were originally carried out by the Research Institute for the Improvement of Costa Rican Education (IIMEC) at the University of Costa Rica, in response to a request from the Ministry of Education for a formative assessment of basic education. From 1988 on, at the request of the Higher Council on Education—the highest authority in Costa Rica’s education system—IIMEC designed and conducted a high school exit examination. The agreement between the ministry and IIMEC expired in 1996, and, the following year, the National Center for Educational Assessment was established as an agency of the Higher Council on Education. The center was dissolved in 1998, and assessment activities are now managed by the Division for the Control and Macro-Assessment of the Education System. This agency of the Ministry of Education administers tests for the formal education system and for the open education subsystem (i.e., the country’s literacy aptitude program, high school distance diploma, and general equivalency diploma). Another office of the Ministry of Education, the Department for Educational Research, carries out studies on the basis of the national test data, although its links to the aforementioned Quality Control Division seem to be very weak.

## **Institutional Framework**

At all levels, including the high school diploma level, tests are referenced to the prevailing national curriculum; to this end, a series of reference syllabuses were devised in developing matrices. Preparation of these syllabuses was performed by the assessment teams; the ministry’s curricular development teams did not participate. Additionally, the standards established by the Central American Educational and Cultural Coordination initiative were not used as referents. The syllabuses are approved each year by the Higher Council on Education, which consults with specialists in curricula for the various subjects before they are approved.

## **Curriculum and Standards**

The tests of cognitive skills were sample-based and norm-referenced. The high school exit exams, as well as the national tests administered at the end of the second and third cycles

## **Types of Tests**

of general basic education, are also norm-referenced but cover all students in those levels of schooling. These latter examinations have been administered on a sample basis (1996–97) and referenced to criteria. All the summative census-based tests (for grade promotion and graduation) account for 60 percent of a student's score in the subjects under assessment. The tests are based on a multiple-choice model, except for the high school graduation examinations, in which an essay composition is also assessed. Efforts were made to include written compositions in the other tests, which were to be graded by designated teachers in each school, but IIMEC research later showed that the grading was deficient; these efforts were discontinued as of 1991.

### **Types of Reports**

Some reports give the national and regional results in criterion-referenced form—that is, in terms of the percentages of students who give the correct answers to the questions in the various objectives of each area of the curriculum. This was the case in the 1996–97 comparative report, which sought to gauge the impact of the corrective measures taken by the Ministry of Education in those years. Other reports, such as that on a sample-based test in 1998, provide the arithmetical mean of achievement by objective in each curriculum area being assessed at the national level; this is then disaggregated by gender, region, school size, and school management (public or private). The reports also provide a descriptive analysis of the results, outline the main difficulties the students faced, and offer a series of methodological recommendations. The official reports on the high school exit examinations present the results in aggregate form at the national level, and disaggregated at the regional level by subject and modality, and provide percentages of students who advance by one grade in each subject. These reports also discuss the results and offer a series of conclusions and recommendations geared to improving education. Studies by the ministry's Quality Control Division provide data on the correlations between test results and student performance in diversified education, as well as information on the predictive value of high school exit examinations with respect to the performance of graduating students in some higher education courses.



The Quality Control Division prepares different reports for the senior personnel at the Ministry of Education, regional authorities and the schools, and parents and students. Technical reports are also developed for researchers, academic specialists, and the media. In addition to the formal reports, in-depth studies of the results are often produced, such as those mentioned above. The high school graduation examinations and those at the end of the second and third cycles of general basic education are high-stakes tests for students, accounting for 60 percent of their individual score. The other 40 percent of their score is from their general average in the last three years of study in each subject. For exiting high school students, the remaining 40 percent is the average of five subjects in their final two years of study. As of this writing, the test results have had little impact on education policymaking at the national level, with the exception of specific changes made to the test syllabuses or to the tests themselves.

There are no subnational assessment systems.

Costa Rica took part in LLECE (1997), but for organizational reasons, the assessment process was not completed, and thus the final results were not published.

Costa Rica's long experience with standardized educational assessment has ensured the acquisition of technical capacity over the years. In general, and despite the continued debate on the utility of national assessment tests, public opinion seems to favor their continuation.

The data provided by the annual tests have not been used for policymaking, nor as inputs for designing and improving programs to raise educational quality. Instead, the data seem to be used almost exclusively to decide whether students should move to the next grade or graduate from school; this raises questions about the tests' formative-evaluative potential.

## **Use and Dissemination of Results**

## **Subnational Systems**

## **International Tests**

## **Progress Observed**

## **Difficulties Observed**

## Cuba

| Test year                                  | Grades tested  | Subjects tested   | Test coverage   | Results reported to   | Results used for |
|--|----------------|---|-----------------|---|------------------|
| Domestic assessments, administered by SECE |                |   |                 |   |                  |
| 1975                                       | 3, 4, 6, 9, 12 | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> </ul> | Regional sample | —   | —                |
| 1996–1998, 2000                            | 6, 9, 12       |   | National sample | <ul style="list-style-type: none"> <li>• Government</li> <li>• Users</li> </ul> | Teacher training |
| 2002                                       |                |   | National census | —   | —                |
| International assessments                  |                |   |                 |   |                  |
| LLECE (1997)                               |                |   |                 |   |                  |

### Institutional Framework

Education in Cuba is measured and assessed through the System for the Assessment of Educational Quality (SECE), a technical unit of the Central Institute of Pedagogical Sciences, under the aegis of the Ministry of Education. The assessment team consists of researchers from the institute and elected representatives of the Directorates of Supervision, Primary Education, Basic Secondary Education, Pre-University Education, and Teacher Training. This group of 20 persons, in turn, receives logistical support from provincial assessment groups and is aided by the coordination efforts of the National Pedagogical University. SECE assessment activities are mainly research-oriented, which is a departure from previous national assessment programs which were geared to accrediting (i.e., promoting and graduating) students in basic and pre-university education.

### Curriculum and Standards

Cuba has a national curriculum for all levels of schooling. The curriculum was changed substantially in 1975 and again in 1987, although it is not considered to have undergone reform as such. Rather, the curriculum was updated periodically on the basis of information provided by the national advisory and oversight system, which is strongly consoli-

dated in Cuba. There are no national curricular standards and, as of this writing, no plans to devise such standards in the short or medium term.

The tests administered by SECE since 1996 are norm-referenced and include both multiple-choice and open-ended questions. They are census-based at the school level and sample-based within each establishment. The results are thus representative at the municipal, provincial, and national levels, but not at the school level because of the size of the sample in each school. In addition to assessing student performance in language and mathematics, the tests provide information on students' values and attitudes. Context questionnaires for teachers, students, and families provide information on the impact of certain non-school factors on learning—such as socioeconomic status, gender, and race/ethnicity. Comparisons are also made between students' academic performance on the standardized tests and achievement in the same knowledge areas as evaluated by the teachers in the schools.

Information unavailable.

The test results have been used mainly in mass teacher training events via national television, and the data inform the continuous advice that schools receive from local and provincial technical teams. The provinces' education management councils are given a report of the results and a guide explaining their interpretation. The data are also used for case studies that shed further light on reasons for unsatisfactory performance in the schools.

There are no subnational assessment systems.

Cuba took part in LLECE (1997).

Officials in the Central Institute of Pedagogical Sciences maintain that there is a policy to integrate the various areas

## **Types of Tests**

## **Types of Reports**

## **Use and Dissemination of Results**

## **Subnational Systems**

## **International Tests**

## **Progress Observed**

of the Ministry of Education. This has made it possible to disseminate information from the assessments, since it is acknowledged that they are critical to more effective education policymaking. Moreover, it is believed that the widespread use of assessment data in teacher training has helped legitimize the standardized assessment system within the institute.

**Difficulties  
Observed**

The teachers and researcher-assessors taking part in SECE initiatives apparently have different perceptions regarding SECE testing and the ways in which it is carried out. For example, even though the external assessment system has acquired legitimacy, teachers do not always feel that the test results faithfully reflect their students' performance. Additionally, despite the fact that assessment results are formally submitted to the Central Subjects Committees—committees of experts responsible for revising the curriculum—there is still a need to ensure that the data are put to better use in improving the curriculum.

**Outlook**

Assessment data should have a greater impact on developing and upgrading the curriculum. It is expected that the external assessment system will acquire greater legitimacy among teachers.

## Dominican Republic

| Test year  | Grades tested  | Subjects tested  | Test coverage   | Results reported to | Results used for   |
|--|--|--|-----------------|---------------------|--|
| <b>Domestic assessments, administered by National Tests System</b> |  |  |                 |                     |  |
| 1991–1996  | 8, 4M, third cycle of adult education                  | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> <li>• Natural sciences</li> <li>• Social sciences</li> </ul> | National census | Users               | —  |
| 1997–2003  | 8, third cycle of adult education, high school seniors |  |                 | Government          | <ul style="list-style-type: none"> <li>• Teacher training</li> <li>• Curricular development</li> </ul> |
| <b>International assessments</b>                                   |  |  |                 |                     |  |
| LLECE (1997)   |  |  |                 |                     |  |

At present, the agency responsible for testing academic performance in the Dominican Republic is the Department of National Testing, which is part of the Education Secretariat's Directorate of Assessment and Control. Originally, national testing was managed and led by an executive committee comprising the secretary of state for education and culture, the rectors of the country's leading universities, and the secretary general of the Dominican Teachers' Association. At that time, the project was financed mainly through international loans from the Inter-American Development Bank and the World Bank. Most assessment programs are now financed through the secretariat's regular budget.

### Institutional Framework

The national tests are based on the content of the national curriculum, and the secretariat's curricular development teams help define the reference matrix and validate the items. Nonetheless, there are usually different viewpoints as to the relevance of the knowledge areas to be tested. Those differences arise because the curriculum does not adequately specify the performance expected in the various areas. Many believe that it is necessary to agree on a set of curricular standards at the national level. The General Directorate for

### Curriculum and Standards

the Curriculum is working along these lines, specifically with regard to Spanish and mathematics.

### **Types of Tests**

Since 1998, norm-referenced tests have been administered in the fourth year of high school and criterion-referenced tests in the eighth grade and third cycle of adult education. High school assessment involves comprehensive tests for all schools and separate tests that distinguish among schools' academic orientation. Both assessments are printed in the same exam book. Tests on the Spanish language assess skills; while those in mathematics and the social and natural sciences assess grasp of content or achievement of objectives. All tests are multiple choice. There are no special instruments to collect contextual data, but information is available on student characteristics (gender, age, and average grades in nationally tested subjects) and school sociodemographics (urban, marginal urban, urban-tourist, rural, isolated rural, area of poverty). The data are used to compile descriptive statistics, as well as establish correlations between the variables mentioned and students' academic achievement.

### **Types of Reports**

Results are reported according to skills attained in a specifications table. School reports present the percentages of students who attained each skill, while individual student reports indicate the performance level in each skill. Institutional reports present the following data disaggregated by course or section and by knowledge area: percentages of students who move forward a grade or are kept behind; correlation between final school score and score in the standardized test; and comparison of results with similar schools and with all schools in the country, region, and district. Student reports present the student's score by domain in each knowledge area, a narrative analysis of main achievements and difficulties by performance level and in comparison to both the national average and the average of schools in the same socioeconomic circumstances or geographic area, and recommendations on improving academic performance. The performance levels were established in line with four categories (quartiles) representing the percentages of students who

could or could not correctly answer questions relating to each skill or subject being tested. Two performance categories are used to discuss results: “low domain” (correctly answering 50 percent of the questions related to the skills or subject) and “domain” (correctly answering 50 percent or more). This information is presented in cross-referenced tables by performance category and level of geographic aggregation (national, regional, district). The executive reports prepared by the Department of National Testing also offer an analysis of the results, as well as a series of conclusions and reflections on their policy implications.

The national tests are used as a partial indicator of academic performance to determine whether students in the levels and courses being tested can move up a grade or graduate. Correct answers to all the test questions account for 30 percent of the pass score for the course, while the maximum score in the last year of the course, together with the average in the last four years of high school, accounts for the other 70 percent. For the eighth grade of basic and adult education, the cut-off for a passing score is 65 out of 100 points; in the fourth year of high school, it is 70 out of 100. Students who fail to reach the minimum passing score in any single subject can take the test again. Students in eighth grade and adult education who do not reach the minimum score in the four subjects have to repeat the course. Final scores are delivered to students as reports (school scorecards), by telephone, or electronically through an official website. The assessment teams have convened other actors in the sector (the curriculum directorate, teacher training institutions, school principals, teachers and students, intermediate authorities, etc.) for critical readings of the results and a discussion of their implications for redefining educational improvement programs and activities. Executive summaries of the results are given to policymakers and technical staff at the national and local levels. Nonetheless, the results are believed to have had little impact on education policymaking in general. The local press publishes the assessment results and comments on them every year, but it is difficult to ensure that the media

### **Use and Dissemination of Results**

engage in substantive discussion of the matter rather than simply presenting overall national averages.

**Subnational  
Systems**

There are no subnational assessment systems.

**International  
Tests**

The Dominican Republic took part in LLECE (1997). Previously, it conducted and published several reports about their application of the Second International Mathematics Study. It also participated in the preparatory phases of curricular analysis of TIMSS, and published their results, but were unable to secure funding for the actual testing.

**Progress  
Observed**

Assessment results are given to the schools and to each student in a way that, from a curricular standpoint, is conducive to educational improvement. Because the tests are governed by the General Education Law and a series of national ordinances, their continuation is guaranteed and they are accorded formal legitimacy.

**Difficulties  
Observed**

Organizational problems in the Education Secretariat hamper the proper flow of information and synergies among the various management units. There is usually an overlap of assessment efforts among the various agencies, and they have only a limited impact on decisions geared to devising comprehensive or complementary policies on quality improvement.

**Outlook**

The National Test Program is expected to continue to operate as it has thus far. From the institutional perspective, the National Education Council recently announced that the National System for the Assessment of Educational Quality would begin operating as a decentralized agency, but the guidelines and specific regulations needed to bring about this change in status had not been drawn up as of this writing.



## Ecuador

| Test year                                     | Grades tested | Subjects tested   | Test coverage   | Results reported to   | Results used for |
|---|---------------|---|-----------------|---|------------------|
| Domestic assessments, administered by APRENDO |               |   |                 |   |                  |
| 1996–2000                                     | 3, 7, 10      | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> </ul> | National sample | <ul style="list-style-type: none"> <li>• Government</li> <li>• Users</li> <li>• Public</li> </ul> | —                |
| International assessments                     |               |   |                 |   |                  |
| None  |               |   |                 |   |                  |

Between 1994 and 2000, APRENDO (Ecuador’s National System for Measuring Academic Achievement) was managed and implemented as part of the World Bank–funded Project for the Development, Efficiency, and Quality of Basic Education. APRENDO was not part of the Ministry of Education’s formal structure, although it provided financial support to the institution, and its activities ended with the expiration of the financial agreement with the World Bank. All the assessment exercises were designed and implemented with technical advice from the Catholic University. Responsibility for assessing education was transferred to the Ministry of Education’s technical staff in 2001; as of this writing, no assessment system has been devised to replace or continue the activities of the APRENDO program.

### Institutional Framework

Ecuador reformed its curriculum and approved new national curricular frameworks in 1996. APRENDO began to define cognitive skills to be assessed before the new official curriculum was approved.

### Curriculum and Standards

Until 2000, APRENDO’s performance tests were sample-based and criterion-referenced. Each student answered four multiple-choice questions per skill. The tests were complemented by questionnaires on the in-school and out-of-school context for students, parents, teachers, and principals. Information is unavailable on the specific variables studied.

### Types of Tests

**Types of Reports**

The overall reports include graphs showing the percentages of students who achieved each of the skills assessed and, in greater detail, each of the “domain” levels attained in each skill. These levels are classified as *beginning* (zero or one correct answer per skill), *progress* (two correct answers), and *domain* (three or four correct answers). The values are presented in aggregate form at the national level, and are disaggregated by geographic region (highlands or coast, and provincial aggregates within each region), school management (public or private), and geographic location (urban or rural). The reports also include a narrative description of the results obtained by various subgroups. Finally, they present a series of general conclusions and describe the implications of the results for devising educational improvement policies. The official reports present neither the findings of the context questionnaires nor any analysis of performance-related factors.

**Use and Dissemination of Results**

The general reports on results were disseminated through the media and distributed to universities, teachers’ unions, and teacher training institutions. The results were also presented publicly in events that were covered by the local media. On the basis of these results, and in line with the substance of the country’s curricular reform, the APRENDO team drew up a proposal on teacher training and presented it to the Ministry of Education; however, the initiative did not make any headway. The team also proposed that the education faculties in several universities change their teacher training curricula on the basis of the APRENDO results; this proposal also failed to bear fruit. The results reports were given to the schools in the test sample, but it is unclear how they used the data. The only available information is on private schools with good results, which used the data for institutional marketing purposes. Until the APRENDO exercises ended, their results had not been used to inform education policymaking at the national level.

**Subnational Systems**

There are no subnational assessment systems.

Ecuador has not taken part in any comparative international tests.

**International Tests**

The APRENDO project made a significant contribution to the development of local assessment capacities and, while the exercises were ongoing, provided data on academic performance that had not been studied previously in Ecuador.

**Progress Observed**

Because the assessment system was not part of the organizational structure of the Ministry of Education but instead was a special international cooperation project, the technical capacities acquired and the data gathered were never properly transferred to the relevant ministry staff.

**Difficulties Observed**

A November 2002 document, *Social Contract for Education in Ecuador*, was published electronically by the Ministry of Education and Culture. This publication announced the government's intention to create a comprehensive assessment system covering not only what students learn but all of the system's mechanisms for management and educational practice. As of this writing, no information is available on whether programs have been devised or implemented to realize these goals.

**Outlook**

## El Salvador

| Test year  | Grades tested     | Subjects tested   | Test coverage   | Results reported to   | Results used for   |
|--|-------------------|---|-----------------|---|--|
| <b>Domestic assessments: SABE</b>                                      |                   |   |                 |   |  |
| 1993–1996  | K, 3, 4, 5, 6, 9* | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> <li>• Natural sciences</li> <li>• Social sciences</li> </ul>  | National sample | Government  | <ul style="list-style-type: none"> <li>• Curricular development</li> <li>• Management training</li> </ul>            |
| <b>Domestic assessments, administered by the Ministry of Education</b> |                   |   |                 |   |  |
| 1998   | 3, 6, 9           | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> <li>• Natural sciences</li> <li>• Social sciences</li> </ul>  | National sample | Government  | <ul style="list-style-type: none"> <li>• Curricular development</li> <li>• Management training</li> </ul>            |
| <b>Domestic assessments, administered by SINEA</b>                     |                   |   |                 |   |  |
| 2001, 2003   | 3, 6, 9           | <ul style="list-style-type: none"> <li>• Environment</li> <li>• Health</li> <li>• Language</li> <li>• Mathematics</li> <li>• Natural sciences</li> <li>• Social sciences</li> </ul> | National sample | <ul style="list-style-type: none"> <li>• Government</li> <li>• Users</li> </ul>                   | —  |
| 2005   |                   | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> </ul>   | National census | <ul style="list-style-type: none"> <li>• Government</li> <li>• Users</li> <li>• Public</li> </ul> |  |
| <b>Domestic assessments: PAES</b>                                      |                   |   |                 |   |  |
| 1997–2005 (annually)   | 2M, 3M**          | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> <li>• Natural sciences</li> <li>• Social sciences</li> </ul>  | National census | <ul style="list-style-type: none"> <li>• Government</li> <li>• Users</li> <li>• Public</li> </ul> | <ul style="list-style-type: none"> <li>• Curricular development</li> <li>• Selection for higher education</li> </ul> |
| <b>International assessments</b>                                       |                   |   |                 |   |  |
| None   |                   |   |                 |   |  |

\*Different grades were tested in different years.

\*\*The test for grade 2M was administered to those in regular high school, while that for grade 3M was for those in vocational education.

The 1991–96 Strengthening of Achievement in Basic Education (SABE) project included a basic education evaluation component, whereby it administered several learning assessments through a contractual agreement with the Intercultural Center for Research in Education. PAES (Learning and Aptitude Test for High School Students) has been administered annually since 1997. It is official and governed by the general law on education; PAES exams are designed and administered by the Central American University José Simeón Cañas in coordination with the Ministry of Education. Since the implementation of the National System for Measuring Learning (SINEA) in 2001, the tests have been based on assessing student competencies. In 2002, the PAES assessment became a part of SINEA, and, since 2005, students' test results have been linked to their ability to graduate (the test counts for 20 percent of the passing grade in each of the main subjects). SINEA has also developed tests for graduates of teaching programs in Ministry of Education–authorized institutions. SINEA is under the authority of the Monitoring and Evaluation Department of the Ministry of Education, and its evaluation activities have been financed with national and foreign resources.

## **Institutional Framework**

The Ministry of Education reformed the curriculum between 1991 and 1997; the reforms were institutionalized under the Ten-Year Plan for Education Reform. The reform affected all grades from kindergarten through high school and was accompanied by in-service teacher training; teacher training programs for future teachers; and the donation of educational materials, including textbooks and libraries, to schools. El Salvador participated in the Central American Educational and Cultural Coordination project to devise its standards. In that context, broad consultations were undertaken with representatives of the teaching sector, the general public, various education-related technical groups, and the business community. Afterwards, it proceeded to establish standards for grades 1 through 6. These standards were tested between 2002 and 2004 in a sample of 110 public schools throughout the country. Based on its experience with curricular reform and standards development and as part of the National

## **Curriculum and Standards**

Education Plan 2021, the ministry updated the curriculum by emphasizing competencies and defining indicators that would measure student achievement on those competencies. The assessment project was developed in tandem with the formulation of the national curricular frameworks. Its reference matrices have gradually incorporated the content prescribed by those frameworks, and the teams responsible for curricular development have played a role in validating the items. The tests were based on a set of universal cognitive skills.

### **Types of Tests**

The first tests were sample-based and multiple choice. A sub-sample of students also took tests in written composition and oral reading. The tests are referenced to assessment criteria consisting of 10 basic performance goals that include specific content for each curricular area being assessed. Since 2001, with the establishment of SINEA, the tests are still criterion-referenced, but they now measure competencies rather than objectives. In 2001 and 2003, tests on the four main academic subjects were administered to a sample of students in grades 3, 6, and 9; the tests included context questionnaires for teachers, principals, and parents. At this time, SINEA experimented with the value-added methodology in three municipalities. For the first time, in 2005, SINEA administered a census-based test to students in basic education on mathematics and language skills. Until 1996, tests were administered by contracted personnel, but since 1998, tests in both elementary and high school education have included the participation of teachers, who are supervised by technical experts from the Ministry of Education. PAES is census-based and has been administered annually since 1997; it evaluates the four main areas of the curriculum. Although it was originally conceived as a norm-referenced test, in 2002 it became criterion-referenced, and it now focuses on evaluating student competencies.

### **Types of Reports**

Up until 1998, the reports presented assessment results in terms of the average percentages of objectives attained by students at the national level and disaggregated by geograph-

ical department, region, gender, and type of school management (public versus private). Under SINEA's administration, the reports show the national average for each of the test subjects and grades and the percentage of students who met the different achievement levels. Each school receives an institutional report that includes scores for each subject and grade, as well as the percentage of students at each achievement level. The census test for basic education in 2005 provided reports to schools with comparative information at the municipal, departmental, and national levels. PAES results are given to each individual student. Every student receives information on the skills assessed in each subject, the extent to which those skills have been attained, the overall performance average in each area, and a conceptual explanation of the achievement levels reached. Reports with comparative information are given to each school.

Generally, until 2000, the results did not affect education policymaking. This is largely because the assessment system was isolated from the ministry's other technical teams for many years, and no effort was made to sensitize members of the educational community or to foster their participation in assessment processes. Central and local users of the system consequently had no demand for the information. Organizational changes were made to the system after 1998 to make it easier for the curriculum teams to take part in validating the instruments and for teachers to participate in administering the tests. The results are given to all evaluated schools, and, in some cases, workshops were held that were intended to develop the skills needed to analyze and compare the data. Teachers' response to the information has been highly positive, but there has not been follow-up to determine if improvements based on the results are being made in the classroom. In addition to the results workshops, a national survey was carried out, which revealed principals' and teachers' concerns about possible reasons for poor performance. Teacher proposals on how to improve matters in each of these areas were collected, and the results of the survey were published in a book by the Ministry of Education. PAES results are given to the students individually. The

### **Use and Dissemination of Results**

results are beginning to be used by universities as a criteria for admission. Furthermore, a minimum score on PAES is required for admission to professional graduate programs. The media widely disseminates student results on the basic education and high school exams.

**Subnational Systems**

There are no subnational assessment systems.

**International Tests**

El Salvador has not taken part in any comparative international tests, but it is scheduled to participate in LLECE (2006) and TIMSS (2007).

**Progress Observed**

Significant progress has been made in strengthening the curricular validity of the assessments, and steps have been taken to disseminate results more widely within the Ministry of Education and among the schools being assessed.

**Difficulties Observed**

For basic education, test administration has not been systematic. Only in 2005 was the first census-based test administered, which has allowed the creation of a baseline that will encourage progress in the schools. Even though results are given to schools and individual reports to high school students, dissemination has not been sufficient, and the impact that the results have at the classroom level could be much greater.

**Outlook**

The results of SINEA-administered tests are expected to be used to change training strategies. Thought has been given to using the data to train teachers and local experts from the Ministry of Education, with an emphasis on those subjects that require reinforcement. It is hoped that the information gathered from the achievement tests becomes part of the National System of Information (which includes other indicators on school coverage, efficiency, and finance) to encourage schools to revise their educational plans periodically.



## Guatemala

| Test year  | Grades tested                  | Subjects tested   | Test coverage                         | Results reported to   | Results used for   |
|--|--------------------------------|---|---------------------------------------|---|--|
| Domestic assessments, administered by SINMELA  |                                |   |                                       |   |  |
| 1992–1996  | 3                              | —   | Experimental sample                   | —   | —  |
| Domestic assessments, administered by PRONERE*                                       |                                |   |                                       |   |  |
| 1998–2000  | 3, 6                           | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> </ul> | National sample                       | <ul style="list-style-type: none"> <li>• Government</li> <li>• Users</li> <li>• Public</li> </ul> | <ul style="list-style-type: none"> <li>• Teacher training</li> <li>• Curricular development</li> </ul> |
| 2004   | 1, 3                           | —   | National sample (public schools only) | —   | —  |
| Domestic assessments, administered by DIGEBI   |                                |   |                                       |   |  |
| 2003   | —                              | —   | National sample                       | —   | Curricular development   |
| Domestic assessments, administered by Ministry of Education/University of San Carlos |                                |   |                                       |   |  |
| 2004   | Final year of secondary school | —   | National sample                       | —   | —  |
| International assessments  |                                |   |                                       |   |  |
| None   |                                |   |                                       |   |  |

\*The 1998–2000 sample included a deliberate overrepresentation of special programs. The 2004 assessment was conducted in conjunction with Juárez Associates.

The first standardized learning tests in Guatemala were carried out at the beginning of the last decade as part of the U.S. Agency for International Development’s (USAID’s) Basic Education Strengthening project, which was coordinated by the National Center for School Performance Tests. Those tests, which were later discontinued, were used to evaluate project implementation in the participating schools. An agreement between the Ministry of Education and the University of Valle de Guatemala led to the creation of the National System for

### Institutional Framework

Measuring Academic Achievement (SINMELA), which was renamed the National Program for School Achievement Assessment (PRONERE) in 1997. This was part of the country's educational reform initiative and received financial support from the World Bank. The university designed and carried out assessment exercises until 2001, when the financial agreement with the World Bank expired. At that point, PRONERE was dismantled and has not been replaced by another assessment program.

### **Curriculum and Standards**

Guatemala lacks a unified national curriculum, and proposals on standards have not made any headway in recent years. The PRONERE tests were originally developed by drawing on an item bank made available by a private educational institution. Later, a reference matrix was developed on the basis of classroom observations and teacher reports on curricular coverage in the classroom. No efforts were made to devise a reference matrix reflecting national agreements on learning priorities to be assessed.

### **Types of Tests**

The PRONERE tests are sample-based, nationally representative, and norm-referenced, and contain 40 questions in each area under assessment. The tests were written and administered in Spanish and four majority indigenous languages. In many bilingual schools, however, the tests could not be administered because students drop out before third grade. The tests were complemented by context questionnaires for students, teachers, and school principals. The tests administered at bilingual schools by the General Directorate of Inter-cultural Bilingual Education are criterion-referenced.

### **Types of Reports**

The PRONERE reports consist of narrative descriptions and comparisons among student subgroups showing the average percentage of correct answers provided for each area being tested. The information is presented in aggregate form at the national level and is disaggregated by department, geographic location (rural or urban), and linguistic group.

Assessment results are used to prepare executive summaries and graphical presentations for national and regional policymakers, as well as for teacher training colleges. The narrative descriptions and pedagogical recommendations are based on a selection of test items that proved hardest to answer. The reports describe the educational establishments being assessed and the socioeconomic level of the families in the sample. They also present some of the contextual factors that have the greatest effect on students' academic performance. These data have been used for some teacher training events, especially those for bilingual teachers through the National Program of Bilingual Education. In general, the results have had little impact on public opinion and education policymaking.

### **Use and Dissemination of Results**

There are no subnational assessment systems.

### **Subnational Systems**

In 2002, the World Bank administered LLECE to a representative sample of third and fourth graders. As of this writing, the results have not been made available.

### **International Tests**

The PRONERE experience made it possible to develop local technical capacities in educational assessment. To the extent that those capacities are used for new undertakings in the assessment field, this is a step forward. Some private sector stakeholders have mobilized to demand that the assessments continue and their results be used effectively.

### **Progress Observed**

The cessation of assessment activities upon expiration of the agreement with the World Bank, and the government's unwillingness to maintain the assessment system, reveal PRONERE's limited degree of institutionalization, despite efforts made to continue the tests for four successive years. Wider dissemination of results and greater participation on the part of key actors in the sector are crucial to rehabilitating and sustaining the assessment system.

### **Difficulties Observed**

As of this writing, the Ministry of Education has made no proposals to reactivate PRONERE or establish any similar agency.

### **Outlook**

## Honduras

| Test year  | Grades tested | Subjects tested  | Test coverage  | Results reported to  | Results used for  |
|--|---------------|--|--|--|---|
| Domestic assessments, administered by Ministry of Education* |               |  |  |  |   |
| 1990–1994  | 1, 2, 3, 4, 5 | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> <li>• Natural sciences</li> <li>• Social sciences</li> </ul> | Experimental sample (10 municipalities)  | <ul style="list-style-type: none"> <li>• Education Secretariat</li> <li>• USAID</li> </ul> | Assessing USAID-supported projects  |
| Domestic assessments, administered by UMCE                   |               |  |  |  |   |
| 1997   | 3, 6          | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> </ul>  | <ul style="list-style-type: none"> <li>• National sample</li> <li>• National census</li> </ul> | <ul style="list-style-type: none"> <li>• Government</li> <li>• Users</li> </ul>            | <ul style="list-style-type: none"> <li>• Teacher training</li> <li>• Targeting support to students and schools</li> </ul> |
| 1998   | 2, 3, 4, 6    |  | National sample  |  | —   |
| 1999   | 2, 3, 4, 5    |  |  |  |   |
| 2000   | 3, 6          |  |  |  |   |
| 2002–2004  | 3, 6          |  |  |  |   |
| International assessments                                    |               |  |  |  |   |
| LLECE (1997)   |               |  |  |  |   |

\*Assessment was undertaken as Component V of the Primary Education Enhancement Project.

### Institutional Framework

Honduras's current assessment system originated in an agreement between the Public Education Secretariat (SEP) and the World Bank regarding the monitoring of the country's Basic Education Improvement Program. Previously (1990–94), efforts had been made to carry out standardized assessments as part of the USAID-financed Primary Education Enhancement Program (PEEP). These programs, however, were never systematized or institutionalized. In 1995,

the Unit for Measuring Educational Quality (UMCE) was created. This unit, which is managed by the National Pedagogical University Francisco Morazán, has a substantial degree of administrative and operational independence, although formally, it reports directly to the SEP's General Directorate for Quality Assessment. The national assessments are coordinated through a decentralized network that operates in the country's 18 departmental directorates; these play a direct role in administering the tests and collecting the data.

Until 2003, Honduras lacked a national curricular framework. Instead, primary-level basic performance and assessment indicators had been drawn up in an ad hoc manner for PEEP, but not all schools had access to the indicators. Because this curricular referent was not appropriate for use in designing an assessment reference matrix, UMCE devised a series of performance objectives and indicators for the various disciplinary areas being assessed. After several years of debate on a variety of curricular proposals, a national curricular framework was officially approved in 2003. Its referents include the content and objectives UMCE had been assessing since 1997. It is unclear how this framework will be integrated and latter assessed with the as-yet uncompleted Central American Educational and Cultural Coordination's standards project.

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.

## Curriculum and Standards

## Types of Tests

UMCE additionally offers its services in assessing academic performance as an indicator of the impact of specific programs financed by international development institutions.

**Types of Reports**

UMCE produced a series of results reports between 1998 and 2001. These reports present performance data in terms of the percentages of students at the national level who meet the objectives in the areas under evaluation. Those objectives are grouped by cognitive domains, domain areas, and specific skills. The data are disaggregated by department, type of school, type of school management (public or private), and geographic location. The reports provide a descriptive analysis of national and departmental results, an examination of selected items, and a series of recommendations for teachers and local technical specialists. UMCE also produces reports on longitudinal studies, multivariate analyses, and comparative analyses of assessment results among grades by subject area, as well as of performance-related factors.

**Use and Dissemination of Results**

UMCE's official reports are submitted to policymakers within SEP and regional educational authorities. It had been hoped that they could also be distributed to the schools, parents, and general public, but various operational and budgetary constraints have hampered such dissemination. Nonetheless, it is reported that about 2,000 teachers, principals, and local technical specialists have had access to the UMCE data since 1998. In many cases, the data come from the technical teams responsible for teacher training. With the support of specialized international organizations, a campaign is currently under way to disseminate assessment results from previous years. Additionally, since 2003, test results have been available on the UMCE website.

**Subnational Systems**

There are no subnational assessment systems.

**International Tests**

Honduras took part in LEECE (1997).

UMCE's internal consolidation has been significant in terms of developing and training technical and professional staff, and in continuing to administer assessments over the years. Efforts have also been devoted to developing a psychometric model of high pedagogical value, particularly in light of the lack of robust and updated curricular referents at the national level.

### **Progress Observed**

The main difficulties in the Honduran assessment system are a lack of communication and weak agreements between UMCE and SEP regarding the targets and uses of the assessments; consequently, the data are underused for policy-making purposes. From a technical standpoint, budgetary constraints have made it impossible to enlarge the sample, which would facilitate hierarchical multivariate analyses of the results and the variables that affect them. These difficulties began to be rectified in 2002 with the introduction of the hierarchical linear model.

### **Difficulties Observed**

It is hoped that assessment results will have a greater impact on public opinion and education policymaking, and that they will be used for teaching and curricular purposes in the schools. Consideration has been given to creating a national assessment system that brings the data from UMCE's external assessment together with that from the internal assessment carried out by SEP itself. It is anticipated that the partnership contract between SEP and UMCE will be extended beyond its current expiration date of 2005.

### **Outlook**

## Mexico

| Test year   | Grades tested          | Subjects tested  | Test coverage   | Results reported to   | Results used for                          |
|---|------------------------|--|-----------------|---|---|
| Domestic assessments: EVEP, administered by SEP and SNEE  |                        |  |                 |   |   |
| 1996–2000   | 1, 2, 3, 4, 5, 6       | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> <li>• Natural sciences</li> <li>• Social sciences</li> </ul>                             | National sample | <ul style="list-style-type: none"> <li>• Government</li> <li>• Users</li> </ul> | Targeting support to students and schools |
| Domestic assessments: Assessment of National Standards  |                        |  |                 |   |   |
| 1997–2003   | 2, 3, 4, 5, 6          | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> </ul>  | National sample | <ul style="list-style-type: none"> <li>• Government</li> <li>• Users</li> </ul> | —   |
| Domestic assessments: Assessment of education achievement component of teaching career program                        |                        |  |                 |   |   |
| 1995–2005   | 3, 4, 5, 6, 1M, 2M, 3M | <ul style="list-style-type: none"> <li>• Foreign language</li> <li>• Language</li> <li>• Mathematics</li> <li>• Natural sciences</li> <li>• Social sciences</li> </ul> | Regional census | <ul style="list-style-type: none"> <li>• Government</li> <li>• Users</li> </ul> | Teaching incentives                       |
| Domestic assessments: IDANIS  |                        |  |                 |   |   |
| 1995–2003   | 6                      | Cognitive skills   | Regional census | <ul style="list-style-type: none"> <li>• Government</li> <li>• Users</li> </ul> | —   |
| International assessments   |                        |  |                 |   |   |
| <ul style="list-style-type: none"> <li>• TIMSS (1995)</li> <li>• LLECE (1997)</li> <li>• PISA (2000, 2003)</li> </ul> |                        |  |                 |   |   |

### Institutional Framework

Until 2002, educational assessment in Mexico was the responsibility of the Public Education Secretariat's General Assessment Directorate. Currently, testing is coordinated between that directorate and the recently created National Institute for Educational Assessment (INEE), a semiautonomous government agency that is part of SEP's operational structure but enjoys greater administrative and political autonomy than other government agencies in the sector. INEE was created following a long debate on the need for a comprehen-



sive assessment system—one that is more objective and efficient, and less vulnerable to circumstantial decisions taken by the authorities in the sector. The education secretariats of all the federal states have their own assessment directorates, although not all of them have their own assessment systems. Those states that have chosen to establish their own systems have taken one of two general approaches: using national tests through expanded or adapted state samples, or devising their own tests and their own means of using and disseminating the results. In both cases, states receive technical support from the General Assessment Directorate and INEE.

The national curriculum in force for primary and secondary education was approved in 1993 as part of a wider reform. The federal government has sole responsibility for setting and updating the curriculum, and state authorities have only limited participation and influence. Because the country already had extensive experience with educational assessment at the time the curriculum was developed, much information on students' educational achievements and difficulties was available. However, the assessment teams had only marginal participation in and influence on curricular decision-making, a situation that persists as of this writing. Although the national tests—except the Instrument for Testing New Secondary School Pupils (IDANIS) and the National University Entrance Exam (EXANI), which assess general cognitive skills—are referenced to the national curriculum, the General Assessment Directorate specified academic content and skills in developing the reference matrices that parallel the official curricular guidelines. Even though this curricular specification, and the tests based on it, are considered “national standards” for primary and secondary education, they cannot really be regarded as such from the perspective of curricular policy. On the other hand, they have helped produce a more complete definition of the educational expectations assessed each year in mathematics and language.

All the national diagnostic tests are sample-based. IDANIS is census-based in the federal district; in the states, it is admin-

## **Curriculum and Standards**

## **Types of Tests**

istered in accordance with specific state requests. All tests are norm-referenced, although the results are reported by performance levels that are determined *ex post*. The tests contain multiple-choice questions and occasionally call for open responses. The census-based teaching career tests are administered to the students of the participating teachers. As an indicator of professional development, these tests are used in determining the teaching career program's incentives scheme. A school self-assessment program has also been developed which is being used by more than 15,000 schools participating in the Quality Schools Program. The self-assessment is adapted from the official Scottish self-assessment model and addresses 33 indicators of school management, teaching practices, and infrastructure. Schools assess their own management on the basis of these indicators and submit their responses to the General Assessment Directorate, which processes the data and prepares summary reports. Additionally, Mexico has secondary school and high school exit examinations to select students who will continue on to further education (high school and higher education, respectively). These examinations, which are known as EXANI I and II, test cognitive skills and specific knowledge in eight disciplines. Regarding the study of performance-related factors, bivariate and multivariate analyses have taken account of in-school variables (such as teachers' knowledge and experience) and out-of-school factors (such as parents' level of schooling and student attendance during preschool). These studies have been complemented by qualitative monitoring of learning outcomes. According to the latest guidelines on assessment policy made public by INEE, the national tests to be administered in the coming years will be census-based in certain key grades and complemented by assessments of teachers, principals, and schools as a whole.

### **Types of Reports**

Executive reports, which are produced for federal officials, present overall results disaggregated by state, geographic location (urban or rural), and percentage of schools that attain each of the established performance levels (very low, low, average, high, and very high). Local supervisors and schools are given reports aimed at improving teaching; how-

ever, such reporting is sporadic, and there is no set reporting format. The IDANIS reports feature graphs showing performance averages at the national and state levels in terms of the arithmetical mean and standard deviation in each case. Schools receive the results of qualitative assessments and research administered by the General Assessment Directorate. These reports describe the institutional characteristics of those schools whose results improve over time, as well as of those whose performance deteriorates. They also suggest how such research can be reproduced in each school through self-assessment and make a series of pedagogical and organizational recommendations for raising educational quality.

Assessment data began to be made public in 2000; previously, results were available only within the Education Secretariat. Of late, several official reports on recent tests have been published, as have summaries of the results at the national and state levels, reports for schools containing data gathered in the self-assessment exercises, and qualitative research carried out by the General Assessment Directorate (see above). As of this writing, the results have been used mostly to determine the allocation of teaching career incentives, and to graduate and select students through the EXANI I and II. INEE has announced its intention to introduce census-based learning assessments gradually in key secondary school grades. These, together with teacher assessments, will help determine whether students advance to the next grade and graduate. INEE's new assessment policy guidelines stress the need for, and commitment to, enhancing the technical transparency of the tests and disseminating the results (including to students and their families), as well as to producing reports for distinct audiences. Within months of its establishment, INEE produced a national results report that was presented to Mexico's president, the media, and state authorities and published on INEE's website.

Several states have set up independent systems to measure and assess educational quality; see p. 148 for a discussion of the system in Aguascalientes.

### **Use and Dissemination of Results**

### **Subnational Systems**

---

|                              |   |
|------------------------------|---|
| <b>International Tests</b>   | Mexico took part in TIMSS (1995), LLECE (1997), and PISA (2000, 2003).  |
| <b>Progress Observed</b>     | Mexico has acquired extensive technical expertise in the area of educational assessment over the past two decades, and there is relative stability among the technical staff dedicated to this work. INEE's establishment, and the long debate that preceded its creation and official approval, underscores the importance that standardized assessment has acquired in public education policymaking. |
| <b>Difficulties Observed</b> | As of this writing, the main difficulty has been that the many assessments carried out have not given rise to a substantive and comprehensive appraisal of educational quality. Because of the lack of communication among SEP units, and the still-deficient mechanisms to disseminate results, data remain unanalyzed and fail to have a greater impact on policymaking.                              |
| <b>Outlook</b>               | INEE's primary challenge is to ensure that the data it gathers are properly analyzed and used to improve the quality of education. From an institutional standpoint, INEE needs to establish its legitimacy within the sector and in the public mind. It can do this by providing useful information to different actors and ensuring appropriate coordination with other government agencies.          |

## Nicaragua

| Test year   | Grades tested | Subjects tested             | Test coverage       | Results reported to                 | Results used for            |
|---|---------------|-----------------------------|---------------------|-------------------------------------|-----------------------------|
| Domestic assessment of primary education curriculum   |               |                             |                     |                                     |                             |
| 1996  | 3, 4          | • Language<br>• Mathematics | Experimental sample | Users                               | Curricular development      |
| 1997  | 5             |                             |                     |                                     |                             |
| Domestic assessments, administered by APRENDE and SNE |               |                             |                     |                                     |                             |
| 1996–1997   | 4, 3M         | • Language<br>• Mathematics | Experimental sample | • Government<br>• Users             | Decentralization assessment |
| 2002  | 3, 6          |                             | National sample     | • Government<br>• Users<br>• Public | —                           |
| International assessments                             |               |                             |                     |                                     |                             |
| None  |               |                             |                     |                                     |                             |

Standardized learning assessment in Nicaragua is carried out by the Ministry of Education’s Directorate for Assessment, Culture, and Sport as part of the Second Project on Basic Education (APRENDE). Until 1999, the directorate’s assessment efforts were relatively uncoordinated and lacked clear goals as to the potential use of the information collected. The National System for Assessing Basic and High School Education (SNE) was established in 1999 and is supported by APRENDE, USAID’s Basic Education Project, and UNESCO’s System for the Improvement of Nicaraguan Education initiative. The new organization is operating in a context of significant reform aimed at decentralizing the administration of public schools, and is reconsidering the country’s overall assessment strategy and objectives.

Information unavailable.

Tests are sample-based, criterion-referenced, and multiple choice. They are complemented by context questionnaires

### Institutional Framework

### Curriculum and Standards Types of Tests

on in-school and out-of-school factors to be completed by teachers, principals, students, and parents. Additionally, there are self-assessment exercises for schools in a small and nonrepresentative national sample.

### **Types of Reports**

Results are reported in terms of percentages of students who are in the low, intermediate, and proficient ranges of the performance scale. Those percentages are presented in aggregate at the national level and disaggregated by region, school management (autonomous, non-autonomous, and private schools with and without a state subsidy), geographic location (urban or rural), and school type (regular or multigrade). The reports also present the results of an analysis of performance-related factors, indicating which in-school and out-of-school variables have the greatest impact on learning.

### **Use and Dissemination of Results**

The Ministry of Education produces technical reports, executive summaries, tables and graphs, and bulletins. There are also workshops, presentations, and colloquies for teachers, principals, municipal and departmental technical specialists and delegates, and senior authorities at the central level. It was announced that results also will be published on the ministry's website, but this had not happened as of this writing.

### **Subnational Systems**

There are no subnational assessment systems.

### **International Tests**

Nicaragua has not taken part in any comparative international tests.

### **Progress Observed**

Efforts have been made—at least in the Directorate for Assessment, Culture, and Sport's formal proposal—to make the national assessment system more significant and useful in education policymaking, as well as in curricular and pedagogical development.

### **Difficulties Observed**

As of this writing, dissemination of results is insufficient to affect public opinion or to spur civil society demands for an improvement in educational quality.

## Panama

| Test year   | Grades tested          | Subjects tested   | Test coverage   | Results reported to | Results used for |
|---|------------------------|---|-----------------|---------------------|------------------|
| Domestic assessments, administered by OAS Assessment Program      |                        |   |                 |                     |                  |
| 1981–1985   | —                      | —   | Regional sample | Government          | —                |
| Domestic assessments, administered by Program of Diagnostic Tests |                        |   |                 |                     |                  |
| 1985–1988   | 6, 6M                  | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> <li>• Social sciences</li> </ul>  | National sample | Government          | —                |
| 1992  | 3, 6M                  | <ul style="list-style-type: none"> <li>• Natural sciences</li> </ul>  |                 |                     |                  |
| Domestic assessments, administered by CECE                        |                        |   |                 |                     |                  |
| 1995  | 1M, 2M, 3M, 4M, 5M, 6M | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> </ul>                             | Regional sample | Government          | —                |
| Domestic assessments, administered by SINECE                      |                        |   |                 |                     |                  |
| 1999–2000   | 3, 6, 9                | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> <li>• Natural sciences</li> </ul> | National sample | Government          | —                |
| 2000  |                        | <ul style="list-style-type: none"> <li>• Social sciences</li> </ul>   | —               |                     |                  |
| 2001  |                        | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> </ul>                             |                 |                     |                  |
| International assessments   |                        |   |                 |                     |                  |
| None  |                        |   |                 |                     |                  |

Several assessment projects have been tried in Panama over the past two decades under different agencies—including the Organization of American States (OAS) and the Center for the Study of Educational Quality (CECE)—but none of them were institutionalized, and the data they gathered were not recorded. The last standardized assessment project in Panama as of this writing was carried out by the Ministry of Education’s Assessment Directorate in 2000 and 2001, but this was a pilot exercise and the sample was small. In

### Institutional Framework

2002, following a wide-ranging intersectoral dialogue, a presidential decree authorized the creation of an assessment directorate with renewed functions and explicit guidelines to develop new and comprehensive assessment exercises and processes.

### **Curriculum and Standards**

The 2000 and 2001 tests were designed while the national curriculum was being updated. The reference matrix was devised by specialists in the disciplinary areas; the Ministry of Education's curriculum teams participated little in the process. A group of specialists from the ministry and from participating educational establishments helped develop the standards promoted by the Central American Educational and Cultural Coordination initiative. The standards in the natural sciences, social sciences, mathematics, and Spanish have been included in the national curricular frameworks for primary education, but nongovernmental sources indicate that they have not been adequately disseminated and have not been complemented by concrete programs for implementation in the schools. The pre-high and high school standards in Spanish, mathematics, the natural sciences, English, and the social sciences are being revised, and will be used to make curricular adjustments in the 7th and 12th grades.

### **Types of Tests**

The 2000 pilot test for sixth graders aimed to determine students' basic skills in problem solving, comprehension, analysis, reflection, and synthesis. Some 18 multiple-choice items and two questions calling for written composition were administered for each subject under assessment. The tests were complemented by a questionnaire on the students' socioeconomic profile, but time constraints made it impossible to process the data for correlation with the final test results. Based on the experience gained through this pilot test, in 2001 the national pilot test *Comprender* was designed for Spanish and mathematics in grades 3, 6, and 9 in basic general schools. The aim was to assess students' verbal and numerical skills and attitudes. These tests also used multiple-choice and composition questions, with a total of 25 questions for each subject being assessed.



The report on the 2000 experience describes the percentage of students who gave the correct answer to each skill being tested, in aggregate form at the national level and disaggregated by provinces and educational regions. It also offers some general conclusions, although these are inadequately developed.

## **Types of Reports**

The results of the 2000–01 tests have not been widely disseminated: reports were given only to regional authorities and the schools in the sample. The data have been underused in the Ministry of Education. The results have informed some training activities, but the main reason for the exercise was to conduct institutional research on attainment levels and the implementation of the new curriculum.

## **Use and Dissemination of Results**

There are no subnational assessment systems.

## **Subnational Systems**

Panama has not taken part in any comparative international tests.

## **International Tests**

The general public favors an appropriate and efficient assessment system.

## **Progress Observed**

The most complex difficulty stems from the limited political will to provide institutional and economic support for the development of a suitable assessment system. Thus far, the Ministry of Education's assessment-related technical agencies have worked in isolation, initiatives are discontinued, and there is a severe outflow of technical/professional resources. Sources consulted recently indicate that efforts are being made to rectify this situation so as to ensure more fluid communication and greater collaboration among the ministry's different players.

## **Difficulties Observed**

Some civic groups, as well as the national legislature's Education Committee, have made important declarations and proposals on national educational assessment and the country's participation in international tests. It is hoped that, in addi-

## **Outlook**

tion to creating a new national assessment directorate in the Ministry of Education, the government will provide institutional support and offer the requisite resources to ensure it functions properly. The new National System for Assessing the Quality of Education (SINECE) will include a system of tests and achievement as well as a system for educational assessment and a qualification and grade advancement system for the first two levels of schooling.

## Paraguay

| Test year                                   | Grades tested             | Subjects tested             | Test coverage    | Results reported to                 | Results used for |
|---|---------------------------|-----------------------------|------------------|-------------------------------------|------------------|
| Domestic assessments, administered by SNEPE |                           |                             |                  |                                     |                  |
| 1996  | 6                         | • Language                  | National sample  | • Government<br>• Users<br>• Public | —                |
| 1997  | 3, 9                      | • Mathematics               |                  |                                     |                  |
| 1998  | 6, 12                     | • Language                  |                  |                                     |                  |
| 1999  | 3, 9                      | • Mathematics               |                  |                                     |                  |
| 2000  | 6, 12                     | • Natural sciences          |                  |                                     |                  |
| 2001  | 3, 6                      | • Language<br>• Mathematics |                  |                                     |                  |
| 2001  | 3, 6                      | —                           | National census* | —                                   |                  |
| 2001  | 3rd year teacher training |                             | National census  |                                     |                  |
| International assessments                   |                           |                             |                  |                                     |                  |
| • LLECE (1997)                              |                           |                             |                  |                                     |                  |

\*Assessed 1,150 schools in the Living School project.

The National System of Evaluation of the Educational Process (SNEPE) is a program managed by the Ministry of Education's Department of Educational Guidance and Assessment. Most of its financing is through the Primary Education Improvement Program by means of an agreement with the Inter-American Development Bank.

### Institutional Framework

The Ministry of Education's curricular development teams take part in the validation of the specifications table, which is referenced to the prevailing national curriculum. At present, there are no proposals on standards, in part because it is believed that establishing excessively high targets would trigger much frustration in most schools.

### Curriculum and Standards

**Types of Tests**

The national-level, sample-based tests are norm-referenced, while the census-based tests in the 1,500 schools participating in the Living School program are criterion-referenced. The tests consist of multiple-choice questions and, for language assessment, written composition. The tests are complemented by student questionnaires aimed at gathering information on attitudes toward the subjects being assessed.

**Types of Reports**

The data are reported in aggregate form at the national level, and disaggregated by geographic department, school management (public or private), geographic location (rural or urban), school size, gender, and session (morning, afternoon, or evening). Results are presented in terms of the arithmetical means and average percentages of correct answers, disaggregated by the skills tested in each curricular area. The information given to the schools in the sample facilitates comparison with counterparts in similar sectors and zones, and with all schools nationwide.

**Use and Dissemination of Results**

Assessment results are provided to the senior decision-makers in the sector and to the schools (principals and teachers) in the form of various reports: one for schools in the sample and another for all schools in the country. The results are presented in regional workshops involving participants from the educational community. Results for a total of 405 schools have also been disseminated through the media, publications, and workshops in various geographic departments. The results of the census-based tests in the Living School program are regarded as one of the main indicators of the program's impact. To that end, comparable tests are administered each year, with a view to gauging how the initiative affects school quality and efficiency after five years. The tests are not high stakes for primary school students, but there are plans to institute accreditation mechanisms for those exiting secondary school through standardized tests over the medium term. Such plans, however, will only be possible after the gradual introduction of the new secondary school curriculum. At present, most efforts focus on delivering results to all schools, so as to demonstrate the system's

transparency and signal its sense of responsibility to the educational community. Assessment results have not been used for specific educational improvement policies and programs, but that remains one of the system's key long-term goals.

There are no subnational assessment systems.

Paraguay took part in LLECE (1997).

Education sector policymakers have remained in place, a positive circumstance for the stability of the technical staff responsible for assessment. Additionally, the evident political will to publicize test results lends transparency to government activities and their impact on education policy. There are also indications that communication among the various agencies in the sector is better than in the past, and, as a result, test data might be used for more informed policymaking in the education sector.

Some segments of the teaching community believe that the content or skills tested are not those that are actually taught or that students are actually ready to learn. More stringent analysis is needed regarding what facilitates and what constrains learning, beyond conclusions based on common sense or speculation. The results reports consist solely of a large number of graphs with brief comments. The same is true of the complementary variables or contextualizing factors: No effort is made to establish a correlation between results and achievements, and thus an opportunity is lost to raise additional questions using other data and more detailed analyses. Government officials do not seem to be considering the establishment of national standards. Current definitions of the domain content assessed are so vague as to be of little use in helping teachers understand what should be improved and how to do so.

SNEPE plans to enhance the dissemination of results within the Ministry of Education so as to ensure that they have a

## **Subnational Systems**

## **International Tests**

## **Progress Observed**

## **Difficulties Observed**

## **Outlook**

greater impact on decisions regarding improvement programs and teacher training.

## Peru

| Test year  | Grades tested | Subjects tested  | Test coverage   | Results reported to   | Results used for |
|--|---------------|--|-----------------|---|------------------|
| Domestic assessments, administered by UMC  |               |  |                 |   |                  |
| 1996   | 4             | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> </ul>  | National sample | —   | —                |
| 1998   | 4, 6, 4M, 5M  | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> <li>• Natural sciences*</li> <li>• Social sciences*</li> </ul> |                 | <ul style="list-style-type: none"> <li>• Government</li> <li>• Users</li> <li>• Public</li> </ul> |                  |
| 2001   | 4, 6, 4M      | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> </ul>  |                 |   |                  |
| 2004   | 2, 6, 3M, 5M  | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> <li>• Civics</li> </ul>  |                 |   |                  |
| International assessments  |               |  |                 |   |                  |
| <ul style="list-style-type: none"> <li>• LLECE (1997)</li> <li>• PISA Plus (2001)</li> </ul> |               |  |                 |   |                  |

\*Administered to grades 4 and 6 only.

External assessment is carried out by the Quality Measuring Unit (UMC), an office of Peru's Ministry of Education. To a significant extent, tests are designed and administered with the participation and advice of external consultants hired by the ministry.

The national curricular frameworks now in force were approved in 1997 and 1998 for primary education, and in 2001 for secondary education (the latter framework is being revised as of this writing). The curricular reforms of the 1990s marked a transition to a skills-based curriculum, and it was on the basis of those skills that the reference matrices for the standardized tests were devised. Designing the matrices, however, called for additional efforts to specify content and put them into effect, since the definition of expected

### Institutional Framework

### Curriculum and Standards

attainment in some areas of the curriculum was too broad. Thus far, there have been no proposals on national curricular standards, although some nongovernmental organizations have offered promising proposals to selected schools.

### **Types of Tests**

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).

### **Types of Reports**

The most recent general reports present assessment results in a criterion-referenced manner, and express the percentage probability that different population groups (such as rural schools, urban public schools, or Quechua-speaking schools) will attain different performance levels in the areas under assessment and their various components or thematic topics. Comparative graphs reveal the performance levels of each stratum in the sample. In-school and out-of-school factors are correlated to overall performance by curricular area, as well as to the various sub-areas being assessed (such as geometry, statistics, inferential text comprehension, etc.).

### **Use and Dissemination of Results**

The dissemination and use of test results are particularly weak aspects of Peru's assessment system. Dissemination of national and international test results was severely limited under the administrations of the 1990s. The data gathered in the 1996 and 1998 tests, as well as those of the UNESCO international test, were made public only in 2000. As of this writing, UMC has published 20 bulletins covering results from the 1998 test, analysis of performance-related factors, analysis of items and incorrect answers on the national



tests for pedagogical purposes, and results of the UNESCO international test. Workshops have been arranged to present results to and discuss them with teachers and specialists from intermediate management agencies. Independent researchers have produced studies on the current curriculum and learning opportunities by analyzing workbooks and exercise books of samples of students in urban, multi-teacher schools. Other studies have focused on performance-related factors and analysis of student compositions. The key challenge in this area is to ensure that official reports are effectively disseminated, that authorities take political responsibility for the results, and that the data are used to improve teaching and inform education policymaking.

There are no subnational assessment systems.

Peru took part in LLECE (1997) and PISA Plus (2001).

Assessment exercises have continued despite changes in technical personnel due to frequent shifts in the sector's management. Results are published after each test, and efforts are being made to produce reports that are more conducive to improving pedagogical management in the schools.

Public officials responsible for education do not make systematic use of assessment data to improve implementation in the most critical areas of the curriculum and in the most disadvantaged schools. This information is also underused in the schools themselves, partly because the results are not distributed to all of them. The technical assessment teams lack the time to develop broader distribution strategies.

The 2003 General Education Law provides for the creation of the Peruvian Institute for Educational Assessment, Accreditation, and Certification. This independent public agency will likely absorb the technology and technical staff currently working in the Ministry of Education's UMC.

### **Subnational Systems**

### **International Tests**

### **Progress Observed**

### **Difficulties Observed**

### **Outlook**

## Uruguay

| Test year   | Grades tested    | Subjects tested  | Test coverage  | Results reported to   | Results used for  |
|---|------------------|--|--|---|---|
| <b>Domestic assessments, administered by UMRE</b> |                  |  |  |   |   |
| 1996  | 6                | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> <li>• Student attitudes</li> </ul>                           | National census  | <ul style="list-style-type: none"> <li>• Government</li> <li>• Users</li> <li>• Public</li> </ul> | <ul style="list-style-type: none"> <li>• Teacher training</li> <li>• Targeting support to students and schools</li> </ul> |
| 1998  | 3                | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> <li>• Natural sciences</li> <li>• Social sciences</li> </ul> | <ul style="list-style-type: none"> <li>• National sample</li> <li>• Self-administered in all schools</li> </ul>                                | <ul style="list-style-type: none"> <li>• Users</li> <li>• Public</li> </ul>                       | Teacher training  |
| 1999  | 6                | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> <li>• Student attitudes</li> </ul>                           | <ul style="list-style-type: none"> <li>• National sample</li> <li>• Experimental sample</li> <li>• Self-administered in all schools</li> </ul> | <ul style="list-style-type: none"> <li>• Government</li> <li>• Users</li> <li>• Public</li> </ul> |   |
| 1999  | 3M               | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> <li>• Natural sciences</li> <li>• Social sciences</li> </ul> | National census  |   | Targeting support to students and schools   |
| 2001  | 4                | —  | —  | —   | —   |
| 2001  | Pre-school, 1, 2 | <ul style="list-style-type: none"> <li>• Cognitive and affective development</li> <li>• Language</li> <li>• Mathematics</li> </ul>         | National sample  | <ul style="list-style-type: none"> <li>• Government</li> <li>• Users</li> </ul>                   | Teacher training  |
| 2002  | 6                | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> </ul>  | <ul style="list-style-type: none"> <li>• National census</li> <li>• National sample</li> </ul>   | —   | —   |
| <b>International assessments</b>                  |                  |  |  |   |   |
| PISA (2003)                                       |                  |  |  |   |   |

Uruguay's Educational Results Measuring Unit (UMRE) was created in 1994 through a financial cooperation agreement between the National Administration for Public Education (ANEP) and the World Bank to develop the Basic Education Quality Improvement Project. Thus, UMRE was not conceived as a national assessment system itself, but as a prototype national system that would lay the foundation for later development of a more comprehensive system to be run from within ANEP. UMRE's assessment activities have been carried out in conjunction with other ANEP administrative units, and, even though it is an agency of the central education apparatus, it has a high degree of technical and administrative independence. Since 2001, UMRE has been institutionalized as part of ANEP's Research and Assessment Management Office, which was created in that year and was then called the Educational Assessment Program. The unit has retained its technical independence but not its administrative autonomy. UMRE is responsible for the assessment of primary and high school education and international testing, and for devising tests to select those admitted to teacher training institutions.

## **Institutional Framework**

Uruguay's national curriculum dates from 1957 but has been amended and updated in recent decades. In the 1970s, it was reformed in line with the behaviorist thinking of the time; its content was highly prescriptive and detailed in terms of coverage. In the 1980s, following the restoration of democracy, a slightly modified version of the 1957 document was developed which proposes a more flexible curriculum which can be adapted to local and institutional conditions. Although the philosophical and epistemological postulates for teaching were amended, the disciplinary content remained the same. The resulting ambiguous curricular reference is thus marked by contradictions between its theoretical basis and the kind of academic content prescribed. This ambiguity posed a significant challenge to UMRE's assessment teams in their efforts to define the specifications table. They resolved the problem satisfactorily, however, through systematic consultation with teachers and experts throughout the country on the relevance and adaptation of the topics and the kind of

## **Curriculum and Standards**

skills that should be assessed. Moreover, one of the main goals of the reference matrices and test model is to help produce a more relevant and updated definition of the national curricular targets. In practice, the conceptual frameworks of the tests, and the free textbooks (with their teaching guides) given to all schools, have become an alternative referent to the national curriculum.

### **Types of Tests**

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).

### **Types of Reports**

Assessment results are grouped by percentages of students who reached the various predefined levels of proficiency (highly satisfactory, satisfactory, unsatisfactory, very unsatisfactory), and are presented in tables that compare performance by geographic department, subsystem (Montevideo, the interior of the country, public and private schools, urban and rural schools), and schools' socioeconomic context. UMRE has emphasized comparison of results across schools from similar socioeconomic backgrounds so as to obviate inferences about the educational quality of schools serving students from different backgrounds. Schools are thus given reports that present their results vis-à-vis schools in the same socioeconomic context within the same geographic department and throughout the country. Overall reports are made public and present results in aggregate form at the national level, as well as disaggregated by department. Data are also disaggregated by school; this information is confidential, and no effort is made to rank schools. Each school receives

its own results and those of similar schools as a whole (all schools in disadvantaged circumstances, for example). Area inspectors are the only other actors with access to school-specific results, and they are expected to use the results to make recommendations on teaching. The presentation of data on proficiency levels is based on a cut-off line (satisfactory level) equivalent to correct answers per student to 60 percent of all items, and on the percentage of students who reach that performance level. The sixth grade reports also provide comparisons over three-year periods of the extent to which the student population has attained the skills being tested. The data are given as percentage points of difference between years 1 and 3 of the test, in aggregate form for all students and areas being tested, and disaggregated by the socioeconomic background of the schools.

The schools are the main audience for the test results, All schools, not just those sampled, receive a report. UMRE technical staff often make presentations and conduct workshops to complement the report. The presentations and published guidance materials focus on analysis of the items and student responses, so as to promote thinking on curricular and pedagogical matters that should be improved. Teachers appear to place a high value on these activities, as evidenced by their requests that UMRE provide more assessment and analysis tools so they can upgrade their teaching practices. UMRE also provides a significant number of test items for teacher use in their courses. Teachers thus have the resources they need to administer their own tests and the information they need to compare their results with those of the national test. This approach is particularly useful when tests are sample-based. Accountability is not a goal of the assessment system. Efforts to legitimize the system in the eyes of the teaching community focus on confidentiality and on using the data solely for pedagogical and curricular purposes. The data have not prompted a public or academic debate, nor responses from civil society, apart from some groups demanding that the results be used to give impetus to accountability and decentralization mechanisms. The data have been used in devising on-the-job teacher training programs, often in coordination

### **Use and Dissemination of Results**

with the primary school inspectors and the Basic Education Quality Improvement Project teacher training program; thus far, however, there has been no impact on initial academic training for teachers. Assessment results were used recently to support the continuation and expansion of an equity program for full-time schools: the data show that such schools attain better results than others in similar socioeconomic circumstances.

### **Subnational Systems**

There are no subnational assessment systems.

### **International Tests**

Uruguay took part in PISA (2003).

### **Progress Observed**

The assessment system enjoys a substantial degree of legitimacy in the eyes of Uruguay's teachers, who view the data as valid and useful inputs for improving teaching practices and school management. The assessment exercises have continued over a period of years, and the technical teams have been stable. Efforts have been made to devise reference matrices and test items that, while consistent with the national curriculum, go beyond what is officially prescribed and set new targets that spur debate on the relevance and validity of what is taught in the schools.

### **Difficulties Observed**

The information disseminated by UMRE has had little impact on public opinion.

### **Outlook**

Assuming that future governments continue to support the country's national assessment system, it is expected that an assessment methodology based on item response theory will be adopted, open-ended questions will be included in sixth grade primary tests, the tests will continue to be administered every three years, and the international PISA test will be used in assessing and monitoring third year high school education and to test those graduating from high school.

## Venezuela

| Test year                                   | Grades tested | Subjects tested   | Test coverage   | Results reported to   | Results used for |
|---|---------------|---|-----------------|---|------------------|
| Domestic assessments, administered by SINEA |               |   |                 |   |                  |
| 1998  | 6             | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> </ul> | National sample | <ul style="list-style-type: none"> <li>• Government</li> <li>• Users</li> <li>• Public</li> </ul> | —                |
| International assessments                   |               |   |                 |   |                  |
| LLECE (1997)                                |               |   |                 |   |                  |

The National System for the Measurement and Assessment of Learning (SINEA) is a program of the Ministry of Education's Vice Ministry of Educational Affairs, although not formally part of the education sector's organizational structure. At the time of the sample-based test in 1998, SINEA was administered by the Sectoral Office of Planning and Budget. Originally financed with World Bank resources, SINEA is now supported by the Ministry of Education's ordinary budget in cooperation with the National Center for the Improvement of Science Education. The 1998 test was carried out with technical support from consultants at the Catholic University of Venezuela, but currently there are few technical and human resources to develop new tests.

### Institutional Framework

The reference matrices informing the design of the tests were developed based on the national curriculum for primary education. As of this writing, there have been no proposals to develop national curricular standards.

### Curriculum and Standards

The 1998 assessment was sample-based and drew on the multiple-choice model. Efforts were also made to use open-ended questions, but the teams responsible for scoring the tests failed to agree on uniform grading criteria. The tests were devised and validated in line with a normative model, but the results were reported in a criterion-referenced fashion, with a cut-off equivalent to 70 percent correct answers.

### Types of Tests

|   |   |
|---|---|
| <b>Types of Reports</b>                 | The 1998 results were reported using cut-off scores on a sliding performance scale. The scale set a level of insufficiency or non-attainment (0–40 points), partial attainment (40–70 points), and sufficiency or attainment (70–100 points). This criterion-referenced report was disaggregated by federal jurisdiction. The results were also presented in a normative manner that showed the performance of different geographic areas relative to the arithmetical mean. The schools in the sample were grouped according to whether they were public or private, urban or rural, and in marginal or less disadvantaged neighborhoods; the results were not reported in line with this categorization, however. The reports for teachers presented the test items and offered a summary analysis of the difficulties students face. They also provided a series of conclusions on the results and their general pedagogical implications. |
| <b>Use and Dissemination of Results</b> | Different reports were produced after the 1998 tests for the central government, for each of the states, and for teachers. These were sent to the education directorates of each state, but it is unclear what real impact they had at the school level. Neither the Ministry of Education nor the intermediate management agencies used the results for policymaking, although the National Center for the Improvement of Science Education did use the results in designing teacher training programs. The media did not formally cover the assessment results, but some articles cited SINEA data (among other indicators) on the poor quality of the public education system.   |
| <b>Subnational Systems</b>              | There are no subnational assessment systems.  |
| <b>International Tests</b>              | Venezuela took part in LLECE (1997).  |
| <b>Progress Observed</b>                | Although it has now been discontinued, the assessment program initially contributed to the development of technical capacities for external assessment.   |



SINEA was insufficiently institutionalized, and its activities ended with the expiration of the agreement with the World Bank. As of this writing, there are not enough human or technical resources to develop or sustain the program, and the Ministry of Education has devoted neither new resources nor management time to this task.

**Difficulties  
Observed**

## Minas Gerais, Brazil

| Test year                                    | Grades tested | Subjects tested   | Test coverage   | Results reported to   | Results used for  |
|--|---------------|---|-----------------|---|---|
| Domestic assessments, administered by SIMAVE |               |   |                 |   |   |
| 2000   | 4, 8, 3M      | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> </ul>             | Regional census | <ul style="list-style-type: none"> <li>• Government</li> <li>• Users</li> <li>• Public</li> </ul> | <ul style="list-style-type: none"> <li>• Teacher training</li> <li>• Targeting support to students and schools</li> </ul> |
| 2001   |               | <ul style="list-style-type: none"> <li>• Natural sciences</li> <li>• Social sciences</li> </ul> |                 |   |   |
| 2002   |               | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> </ul>             |                 |   |   |
| 2003   |               | <ul style="list-style-type: none"> <li>• Language</li> </ul>                                    |                 |   |   |
| International assessments                    |               |   |                 |   |   |
| None   |               |   |                 |   |   |

### Institutional Framework

The Minas Gerais Educational Evaluation System (SIMAVE) is one of the central components of the Assessment Program of the Basic Public Education Network, and is part of the Minas Gerais Education Secretariat. The assessment system was created as part of a broad-scale review of education policies when new authorities assumed responsibility for the sector in 1999. Inequity in access to education within the state, in addition to multiple problems of students repeating years and automatic grade promotion, required a diagnostic assessment system that would enable support to be targeted to these concerns. Testing is one component in a set of programs aimed at supporting teachers and initial teacher training. One of the system's most novel features is that it entails formal collaboration with 29 of the state's higher education institutions. The collaboration involves designing and administering the tests, processing the data, and delivering the results to the schools. The tests are coordinated by one university in collaboration with the others. This process aims to foster the decentralization of educational management as a whole and encourage greater participation by local actors. The partnership is also motivated by the need to provide more information to teacher training institutions and gradu-

ally integrate a stronger assessment culture within the system. SIMAVE collaborates with the SAEB exercises (see Brazil, p. 64) and draws on the national item bank in devising its own tests, but it does not use the national data to formulate state policies.

The state's main curricular referent is the national curricular parameters, although the state education secretariat has a curricular proposal of its own that is consistent with those parameters. The schools develop their own curricular proposals on the basis of these two referents, but the process is a very slow one. SIMAVE's reference matrices are prepared using the national curricular parameters, the state's own curricular proposal, and the SAEB content matrix. The challenge of prioritizing the most important knowledge areas for each grade is undertaken by the technical assessment staff; the teams responsible for curricular development take little part in this initiative. The latter have demanded that SIMAVE be more open, so that the tests and their content are more relevant and legitimate.

The SIMAVE tests are census-based and strictly normative, since they are essentially designed to compare the performance of different groups of students. Efforts have been made to report results in a criterion-referenced fashion, disaggregating data by the kinds of skills attained in each curricular area being tested. Questionnaires are used to gather data on schools' socioeconomic characteristics, but there is no statistical analysis of performance-related factors.

Every school receives an individual report. These offer separate graphs for each grade and curricular area tested, each of which shows the percentage of students in each of the three predetermined performance (proficiency) levels—critical, intermediate, and sufficient. The percentages are disaggregated by the different abilities that comprise each of the skills being tested in the particular curricular area. The reports also include a bar graph comparing the school's average performance relative to the corresponding municipality,

## **Curriculum and Standards**

## **Types of Tests**

## **Types of Reports**

the regional superintendency, and the state. The reports are complemented by an explanatory review to help teachers understand and interpret the data.

### **Use and Dissemination of Results**

The results reports have two main audiences: regional authorities and school staff (principals and teachers). The former are given reports on the overall performance of different regions, superintendencies, and municipalities with a view to helping them devise policies and target materials and teaching support to those regions and superintendencies where performance is poor. Schools receive the reports described above. SIMAVE has pursued legitimization through a diagnostic assessment model that does not imply sanctions or rewards for local actors but instead serves mainly to guide and enhance pedagogical and curricular processes. Another prime goal is to deliver test data to those studying education, and to future teachers, by involving the universities in the entire assessment process (university students are often hired to administer the tests). Consequently, it is expected that future generations of teachers will have greater collective participation in substantive education issues. The results are made available to the general public and the media. Significant efforts are made to ensure that the media do not use the data in a sensationalistic way, especially by publishing rankings and making potentially unfair comparisons. Parents do not receive the results individually; instead, results are given to parent representatives on the governing boards of each school. These representatives have access to the reports and can decide whether to make the information available to other parents. The data have been used to devise teacher training programs, particularly to address the reading problems uncovered; specific school support programs have been designed to that end. The data also underpin the selection of schools that will be accorded priority attention through such programs. Those schools are subject to qualitative research into the reasons for poor performance.

### **International Tests**

Minas Gerais has not taken part in any comparative international tests independent of the national sample.

Despite limited communication between the assessment teams and other offices of the secretariat, the curricular development staff is beginning to use the test data to propose new approaches within their field. As of this writing, attention has focused mainly on the reading problems uncovered. The relationship between SIMAVE and the universities seems to be beneficial, and both parties have shown a willingness to continue the partnership.

**Progress  
Observed**

Some critics argue that teachers should be more directly involved in the assessment process—especially in validating the content tested by SIMAVE—if the tests are to have greater impact. This lack of participation is partially due to the fact that teachers do not feel responsible for the results and do not encourage their students to improve their performance.

**Difficulties  
Observed**

This system was only recently implemented; it is expected that the assessment exercises will become institutionalized and will continue. It is also expected that they will influence teacher training programs.

**Outlook**

## Paraná, Brazil

| Test year  | Grades tested | Subjects tested  | Test coverage   | Results reported to   | Results used for   |
|--|---------------|--|-----------------|---|--|
| Domestic assessments, administered by Paraná's Education System Assessment Program |               |  |                 |   |  |
| 1995   | 4             | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> </ul>  | Regional census | <ul style="list-style-type: none"> <li>• Government</li> <li>• Users</li> <li>• Public</li> </ul> | <ul style="list-style-type: none"> <li>• Teacher training</li> <li>• Curricular development</li> </ul> |
| 1996   | 8, 2M         | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> <li>• Natural sciences</li> <li>• Social sciences</li> </ul> |                 |   |  |
| 1997   | 4, 8          | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> <li>• Natural sciences</li> </ul>                            |                 |   |  |
| 1998   | 4, 8          | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> <li>• Natural sciences</li> </ul>                            |                 |   |  |
| 2000   | 4, 8          | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> <li>• Natural sciences</li> </ul>                            |                 |   |  |
| 2002   | 4, 8, 3M      | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> <li>• Natural sciences</li> </ul>                            |                 |   |  |
| International assessments  |               |  |                 |   |  |
| None   |               |  |                 |   |  |

### Institutional Framework

The Paraná Education System Assessment Program is part of the state's Education Secretariat and comes under the aegis of a broader initiative known as the Teaching Quality Project, which began in 1994 with World Bank support. The Bank's monitoring requirements underlay the design of the system's first assessment instruments. Among these, assessing learning (and the continuation of such assessment) became particularly important. The program has little autonomy, since all technical and policy decisions are made at higher levels of the Education Secretariat. In addition to assessing

the implementation and impact of specific educational quality programs—such as accelerated learning and funding for libraries, teaching materials, and teacher training—the state needed an in-depth appraisal of two key problems thought to be strongly linked to academic performance: grade repetition and over-age students, and dropping out of school. Moreover, the Education Secretariat needed quality indicators to facilitate the monitoring of school decentralization and to target support to the neediest schools. The state has technical exchanges with SAEB (see Brazil, p. 64), especially with regard to curricular matrices and assessment methodologies, but Paraná does not use the data collected by the national assessment agency.

The tests' curricular matrices are referenced to the state's Basic Public Education Curriculum and are aligned with the national curricular parameters. The secretariat's curricular staff helped develop these parameters; they also participate in devising the state's curricular assessment matrices. Until the mid-1990s, teachers were guided essentially by commercial textbooks and were unfamiliar with either national or state curricular parameters. Through the assessments and the training initiatives designed on the basis of assessment results, the secretariat seeks to reverse this trend. The curricular teams maintain that the contents of the matrices are duly validated in terms of their alignment with the state's curricular proposal, although they question how exhaustively the matrix allows students' capacities to be measured.

The tests are norm-referenced and are grounded in classical theory and item response theory. The tests are census-based (in all schools with more than 50 students in the grades being tested), and each student answers 30 questions in each curricular area under assessment. Questionnaires are used to gather information from teachers and principals in order to analyze performance-related factors. The main such factors examined are school management and administration, and the economic and social circumstances of students and their families. Local education teams and teaching staff are exten-

## **Curriculum and Standards**

## **Types of Tests**

sively involved in test administration, as are those parents designated to monitor assessment processes and disseminate information on the results.

### **Types of Reports**

The data given to the schools consist of simple frequencies and central tendency measures of correct answers, aggregated at the state level and disaggregated by municipality, school group, school, and age bands within each group of students taking the test. The reports also present results in terms of percentages of students who attained each of the five levels of difficulty on the model's performance scale. Additionally, the reports specify which of the 30 questions the students answered in each area corresponding to those performance levels. With these data, and following guidelines and formats provided by the secretariat, schools prepare their institutional reports on student performance, identify problems, and propose pedagogically viable courses of action. In the other general reports produced by the assessment system, data are disaggregated by municipality and by other classifications such as public/private school, state/municipal school, and daytime/nighttime classes. They offer a summary of the most critical curricular content (in terms of poor performance) and a pedagogical analysis of some of the test items in order to guide teachers in their planning and in resolving any learning problems uncovered by the assessment. The open-ended questions (written composition) are similarly analyzed to provide clearer indicators of the kind of performance expected of students. Additionally, the Education Secretariat provides test results in the form of school bulletins. These offer data on the performance of each school's students in line with the performance scale, as well as relative to other schools in the municipality and throughout the state. They also present the findings of the surveys used to elicit parent and student opinions on various aspects of school life in each institution. A second set of complementary information in these bulletins includes institutional data on pass, fail, and dropout rates, as well as teacher profile information. All of this can be compared with corresponding information on other schools in the municipality and the state.



The system stresses schools' use of the data through the mechanisms mentioned above. Additionally, the Education Secretariat uses the information in all of the state's training events. Such events, which are usually located so as to bring together teachers from several municipalities, are attended by parent representatives for each school. The aim is to develop an assessment culture on the basis of parent requests for information. The data's use in policymaking is relatively limited, even though this was one of the main goals when the system was first established; the chief reason for this discrepancy is that the assessment teams have little time to prepare more detailed analyses of the data collected. The databases can be accessed by individuals in academia and external researchers, following a formal request to the Education Secretariat. In general, there is little demand for these data. Information is also made available to the media, but it is difficult to prevent journalists from reporting it in a simplistic and sensationalistic way. According to assessment system authorities, the data have little impact on public opinion.

Paraná has not taken part in any comparative international tests independent of the national sample.

Several concrete activities have been undertaken to ensure that the data are used effectively and that the central and local users of the system have real ownership of it. The technical teams are relatively stable, largely because of the permanence of the political authorities in the education sector during recent administrations. There is good institutional collaboration among the teams responsible for assessment, teacher training, and curricular development.

As of this writing, the annual administration of the tests has left little time for technical specialists to undertake more detailed analyses of the data on school performance.

There are plans to administer the assessment tests less frequently, perhaps every two years, so as to allow more time for data analysis and dissemination.

## **Use and Dissemination of Results**

## **International Tests**

## **Progress Observed**

## **Difficulties Observed**

## **Outlook**

## São Paulo, Brazil

| Test year                                    | Grades tested | Subjects tested  | Test coverage   | Results reported to   | Results used for   |
|--|---------------|--|-----------------|---|--|
| Domestic assessments, administered by SARESP |               |  |                 |   |  |
| 1996   | 3, 7          | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> <li>• Natural sciences*</li> <li>• Social sciences*</li> </ul> | Regional census | <ul style="list-style-type: none"> <li>• Government</li> <li>• Users</li> <li>• Public</li> </ul> | <ul style="list-style-type: none"> <li>• Teacher training</li> <li>• Curricular development</li> <li>• Management training</li> <li>• Targeting support to students and schools</li> </ul> |
| 1997   | 4, 8          | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> <li>• Natural sciences*</li> <li>• Social sciences*</li> </ul> |                 |   |  |
| 1998   | 5, 1M         | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> <li>• Natural sciences</li> <li>• Social sciences</li> </ul>   |                 |   |  |
| 2000   | 5, 7, 3M      | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> <li>• Natural sciences</li> </ul>                              |                 |   |  |
| 2001   | 4, 8          | <ul style="list-style-type: none"> <li>• Language</li> </ul>   |                 |   |  |
| International assessments                    |               |  |                 |   |  |
| None   |               |  |                 |   |  |

\*Administered to grades 7 and 8 only.

### Institutional Framework

SARESP—the System for the Assessment of School Performance in São Paulo—is managed jointly by the state Education Secretariat and the Foundation for Educational Development (FDE), an agency of the secretariat that has a significant degree of administrative independence. FDE receives financing from the secretariat but also funds itself by selling services to other agencies. This partial financial autonomy makes FDE more flexible and stable in providing assessment services to the state. However, all decisions regarding assessment coverage and use of assessment data are made primarily by the Education Secretariat. The state's

education authorities decided to create their own assessment system in order to resolve two critical problems: equity and school autonomy. Test data are used to target support to schools whose performance is most critical (equity), and the assessment process itself involves the permanent participation of the schools (which are being decentralized) to ensure that the data are used for pedagogical and administrative purposes (school autonomy). The assessment system is also intended to undertake longitudinal appraisals of the impact of state education policies. SARESP collaborates with the administration of the national SAEB tests (see Brazil, p. 64) but its instruments and assessment exercises are wholly independent. Technical support for these activities (psychometric model, statistical processing, and production of reports) was provided from the outset by the Carlos Chagas Foundation. At present, the latter is also cooperating, through training initiatives, in transferring SARESP's technical coordination to the state Education Secretariat.

The main curricular referents for devising the assessment instruments are the national curricular parameters and the curricular and pedagogical proposal prepared by the state's Education Secretariat. The reference matrix is developed in broad consultation with the Coordinating Office of Teaching Studies and Norms (CENP), a key unit within the secretariat.

The system uses norm-referenced tests based on a set of defined skills in each curricular area to be assessed. The design process involves defining a five-level achievement scale for each area. Each student's performance is classified according to these levels, as are those of the various subgroups and the group in the aggregate. The tests were originally sample-based, but it was decided to move to a census-based model in order to foster accountability for results among the schools. The tests are administered in the final grade of each four-year school cycle. Questionnaires for principals, supervisors, and teachers collect information on school characteristics. This information is used to analyze performance-related factors such as child labor; student

## **Curriculum and Standards**

## **Types of Tests**

ethnicity, age, socioeconomic status, gender, and school trajectory; characteristics of the teaching staff; school environment; and school participation in special programs. Additionally, parent observers answer a list of questions on how the test was administered and whether the process was transparent. The data are processed by the schools themselves in line with procedures established by the central authorities; they are then submitted to the Education Secretariat. Control mechanisms are in place to monitor the accuracy and general consistency of school-level data.

### **Types of Reports**

The reports prepared by the schools present the percentage of correct answers (and the performance level corresponding to that percentage) for all students tested. The school submits its institutional results to the secretariat but keeps the student-specific data. FDE produces reports on the relative and comparative performance of the various subgroups, as well as on the performance-related factors for the group as a whole. The data are disaggregated by student age, type of school, student ethnicity, school timetable (daytime or nighttime), and other factors. The results from the 1996 and 1997 tests were presented in a comparative form. Comparability was assured by using anchor test items—standard questions that are applied from one test to the next—in each test.

### **Use and Dissemination of Results**

SARESP tests are both low and high stakes. The data are used mainly to inform decisions on pedagogical and curricular policy, in the form of special programs that provide support and other inputs to schools with poor results (e.g., support and makeup programs during the year and over the vacation period). The results are high stakes for teachers and students. For teachers, the SARESP results are one of several performance indicators (primarily including retention rate); they might receive a bonus based on their school's overall results. Student advancement to the next school cycle depends on their performance in the annual tests. Deficient performance on the test (less than 50 percent correct answers) can be revised at the discretion of administrators if student performance was very good for the rest of that cycle; in such cases,

students can take the test a second time. As of this writing, the use of assessments in allocating bonuses and advancing students is being reconsidered. Test results are disseminated to the general public via media reports, but presenting rankings is discouraged. The Carlos Chagas Foundation has analyzed the results independently for research purposes, since the databases are accessible to any non-official user. Education policy measures designed on the basis of the SARESP data include a training program that the Education Secretariat has delegated to three universities (São Paulo State University, Campinas State University, and São Paulo University). Training is offered for schools that so request it because of unsatisfactory results on the state tests. This training has focused on developing reading comprehension skills, among other items. The results are also used to assess the impact of special secretariat programs, especially the Accelerated Learning Program. CENP officials believe that progress made in school learning, and documented through SARESP, may be ascribed to a series of measures and programs promoted by the state. There are no qualitative monitoring factors that could explain the test results.

São Paulo has not taken part in any comparative international tests independent of the national sample.

**International  
Tests**

A salient feature of the system is its organizational integration with other departments of the educational administration, especially with those responsible for pedagogical norms who inform the design of support and makeup programs. The substantial participation of the local education community throughout the assessment process is beneficial.

**Progress  
Observed**

It is expected that the assessment system will continue to be consolidated and that the technical assessment capacities of the professional staff in the Education Secretariat will be strengthened.

**Difficulties  
Observed**

## Bogotá, Colombia

| Test year  | Grades tested | Subjects tested  | Test coverage   | Results reported to   | Results used for   |
|--|---------------|--|-----------------|---|--|
| Domestic assessments, administered by Subdirectorate of Assessment and Analysis* |               |  |                 |   |  |
| 1998 <sup>A</sup>  | 3, 5          | <ul style="list-style-type: none"> <li>• Civics/values</li> <li>• Language</li> <li>• Mathematics</li> <li>• Natural sciences</li> </ul> | Regional census | <ul style="list-style-type: none"> <li>• Government</li> <li>• Users</li> <li>• Public</li> </ul> | <ul style="list-style-type: none"> <li>• Teacher training</li> <li>• Curricular development</li> <li>• Management training</li> <li>• Targeting support to students and schools</li> </ul> |
| 1999 <sup>B</sup>  | 3, 5, 7, 9    |  |                 |   |  |
| 1999 <sup>A</sup>  | 7, 9          |  |                 |   |  |
| 2000 <sup>A</sup>  | 3, 5          |  |                 |   |  |
| 2001 <sup>B</sup>  | 3, 5, 7, 9    |  |                 |   |  |
| 2001 <sup>A</sup>  | 7, 9          |  |                 |   |  |
| 2002 <sup>A</sup>  | 3, 5          |  |                 |   |  |
| 2003 <sup>B</sup>  | 3, 5, 7, 9    |  |                 |   |  |
| International assessments  |               |  |                 |   |  |
| None   |               |  |                 |   |  |

Note: Schools in Bogotá follow one of two calendars. Calendar A schools include most public schools and private non-elite schools; Calendar B schools, which follow a Northern hemisphere academic calendar (September–May) are elite private schools and a few public schools. Testing in Calendar A schools is indicated with <sup>A</sup>; in Calendar B schools with <sup>B</sup>.

\*Basic competency tests.

### Institutional Framework

The assessment system in Colombia's capital district is coordinated by the Directorate for Assessment and Support, an agency of the academic undersecretariat of Bogotá's Education Secretariat. The Subdirectorate of Assessment and Analysis and the Subdirectorate of Educational Improvement collaborate under this directorate. The Basic Skill Assessment program was designed and operationalized as part of a significant restructuring of educational administration, and test results are seen as key inputs for decisions geared to fulfilling the Sectoral Management Plans 1998–2001 and 2001–04.

### Curriculum and Standards

Schools in the capital district ascribe to national curricular frameworks (see Colombia, p. 75). The local Education Secretariat can choose to define a series of basic skills to clarify

expectations in four areas. This task was carried out by academic specialists, several of whom had helped develop the national curricular frameworks in previous years. These skills, whose disciplinary content is consistent with the national curricular guidelines and achievement indicators, comprise the reference matrix for designing new standardized assessment instruments. Technically and politically, the basic skills proposed can be regarded as curricular standards. They provide a clear operational description of possible performance levels for each content item assessed, and offer an effective means by which key stakeholders can discuss what students in the system are expected to learn.

The performance tests are criterion-referenced, and are census-based in all the grades being assessed. They include multiple-choice items and open-ended questions. Questions in the area of civic understanding and awareness focus on issues of moral development, social representation, democratic attitudes, and knowledge of the Colombian state.

The Education Secretariat distributes results reports to every school. In these, results for each subject are disaggregated by skill and by performance levels in each skill; the reports present the average value students attain on a scale of 0 to 306. In the general reports made publicly available, the average percentages of achievement are disaggregated by subject, school management (public or private), and geographic area within the district. The reports also provide examples of the test items, an analysis of these, and a series of conclusions on the overall results of each test.

The test results have been used for various purposes, mainly for training teachers and principals and for targeting support to schools. The training materials and events focus on developing the skills needed to analyze the data and on designing pedagogical and curricular improvement strategies on the basis of the results; they make use of printed material, videos, DVDs, and interviews with teachers. They provide specific recommendations, which are presented as

### **Types of Tests**

### **Types of Reports**

### **Use and Dissemination of Results**

possible “improvement paths” that can be used by schools as a model of curricular design and pedagogic planning. In targeting support to specific schools, the Education Secretariat develops projects that seek to take ideas and educational management tools from institutions with better results for replication in other schools where the need for improvement is most apparent. Schools with poor results can also access other support programs, such as cultural outings in the city or the new network of libraries inaugurated in the last five years. The test results and other indicators are being used to inform decisions on investments in infrastructure, facilities, and institutional reorganization. Political accountability for results is assured by the Sectoral Plan 2001–2004, which specifies that the target result is a minimum average of 180 points (on a 0–306 scale) and that all schools should surpass 100 points on the same scale. These targets have been made public, and the local authorities must answer for the results at the end of the period. Schools are encouraged to inform their community of the progress they have made toward attaining the skills being tested. They do this through prominently placed bulletin board displays which are updated with the annual results. Additionally, student grades, which are sent to students and their families in periodic bulletins, are presented in line with the same set of basic skills assessed by the Education Secretariat in standardized form. In conjunction with other indicators, the results are used to monitor the quality of private schools that accommodate public enrollment (a strategy used by the secretariat to guarantee access to schooling while new public schools are being built). The data are also used as one of the indicators of teaching performance in the new system for regulating tenure.

**International Tests**

Bogotá has not taken part in any comparative international tests independent of the national sample.

**Progress Observed**

The most prominent advances include the development of clear and operational curricular standards to assess learning, efforts made to institutionalize and legitimize the assessment system and the information produced, and the stability



of the technical teams and assessment exercises over the past five years. Also noteworthy is the use of the test data in improving teaching/learning processes and for education policymaking in general.

## Aguascalientes, Mexico

| Test year   | Grades tested | Subjects tested   | Test coverage   | Results reported to   | Results used for   |
|---|---------------|---|-----------------|---|--|
| Domestic assessments, administered by State Assessment System |               |   |                 |   |  |
| 2000–2003   | 6             | <ul style="list-style-type: none"> <li>• Language</li> <li>• Mathematics</li> </ul> | Regional census | <ul style="list-style-type: none"> <li>• Government</li> <li>• Users</li> <li>• Public</li> </ul> | <ul style="list-style-type: none"> <li>• Teacher training</li> <li>• Curricular development</li> <li>• Management training</li> <li>• Targeting support to students and schools</li> </ul> |
| International assessments                                     |               |   |                 |   |  |
| None  |               |   |                 |   |  |

### Institutional Framework

Educational assessments in Aguascalientes are managed by the State Assessment System, a technical unit of the state Directorate of Training, Upgrading, and Continuing Professional Education for the Teaching Profession. Assessment activities respond to the priorities of education policy as set out in the State Development Plan 1998–2004, and more specifically in the Educational Development Program 1999–2004. The educational testing and assessment program is part of a process of restructuring the school supervisory system, especially the redistribution of professionals according to the demands and needs of each locality. The aim is to make the system less bureaucratic and administratively complex, and more effective in performing its pedagogical functions. The restructuring was politically and administratively difficult, but made it possible for test data to be used effectively later. A general pedagogical coordinator, who reports directly to the state's secretary of education, ensures that work is coordinated among the directors of different management units, and between those and the local school supervisors. At present, all improvement programs focus on academic quality in the schools. This approach stems from the conviction that student socioeconomic level is not the main determinant of educational attainment, and that appropriate pedagogical policies can make a greater difference to learning than circumstances in the schools.

Like the other states in Mexico, Aguascalientes uses the national curriculum as the main reference for teaching. The reference matrix for designing new assessment instruments is based on the curriculum, and the test questions are taken from the item bank made available by the national Public Education Secretariat's General Directorate for Assessment. The curricular content tested is not validated at the state level, but the items chosen for the tests are subject to local validation—as are the instruments—through consultation with local experts. The authorities believe that content validation is not a crucial element of the assessment system, since the national reference curriculum is accepted by the teachers, but that it is important to legitimize the content by ensuring that the process is transparent and technically appropriate. There are no state-level standards, but the assessment targets for the tests are themselves an indicator of the “official” expectation of the minimum to be attained by all sixth grade students.

The state tests are criterion-referenced and census-based in the sixth grade of primary school. They are based on a model of progress or “educational gain.” This model involves a diagnostic test at the start of the academic year and a summative exam at the end. Comparison of the two sets of results produces an index of the progress made by students in the course of the year. From a curricular standpoint, the model is viable, since the content in fifth and sixth grade language and mathematics is very similar and differs only in depth. The tests include multiple-choice items, and questions that call for written composition are now being included. The aim is to discern the school's capacity to intervene in the development of writing skills. A scale was drawn up with 20 aspects to be evaluated in each written response. The writing will be graded by school supervisors, principals, and teachers, who are now being trained on how to do this in line with uniform criteria. The learning data are complemented by questionnaires on the in-school and out-of-school context for students, teachers, and principals. There are also federal examinations (IDANIS, EXANI I and II), national standards for primary and secondary education, and the Scottish model of

## **Curriculum and Standards**

## **Types of Tests**

school self-assessment. All these instruments, however, are centrally designed in the SEP; Aguascalientes simply administers the tests and uses the information collected. Though coordinated by the SEP, the school's self-assessment program involves teams of local researchers who decide which schools (usually those with the poorest results in the state tests) will derive the most benefit from participating in the program.

### **Types of Reports**

The state test results are reported by performance levels attained in each area. The levels correspond to a four-way division on a scale of 1 to 100 and are classified as critical and insufficient (non-domain) and acceptable and desirable (domain). The year-end educational gain is reported on a scale of -4 to 4, depending on the progress made between the administration of the initial and final tests. For instance, if a student began the year at an insufficient level and ended it at a desirable level, the gain—according to the formula used—is calculated at 3 out of 4 points. In the opposite case—if the student began at a desirable level and performance deteriorated to insufficient—the gain is expressed in negative terms, -3 points on the scale. These data are presented in aggregate form at the state level and disaggregated by locality, school, class, and individual student.

### **Use and Dissemination of Results**

Assessment results are submitted to the senior levels of government and the Education Secretariat, coordinators and supervisors of local education districts, principals, teachers, students, and families. The schools receive a report on performance levels and educational gain, disaggregated by course and individual student. For each group and student tested, the performance levels in both areas under assessment are disaggregated into 50 items of knowledge. In this way, the teacher can see exactly which curricular targets each student met and to what extent the student did so. Supervisors receive information disaggregated by school and class. The reports also present the findings of the context questionnaires in terms of percentages of answers to the chosen items in each group (for example, on the level of parents'

schooling or attitudes to the subject); no effort is made to establish correlations or draw inferences. The Education Secretariat, in collaboration with the district supervisors, undertakes continuous monitoring of the levels of educational gain or progress in the different schools, and asks principals to make recommendations on how to overcome the problems revealed. The secretary of education personally visits the schools, without warning, to verify that the principal has read the reports and is familiar with the difficulties revealed by the tests. The improvement projects the schools are asked to implement call for the use of the available statistics and analysis of the in-school and out-of-school factors that have been shown to affect performance. Emphasis is placed on the in-school factors, since it is here that schools can have a direct effect by devising new teaching and curricular strategies. The improvement initiatives the schools suggest must specify the kind of quantitative progress expected for the following year, typically in the form of an increase in the percentage of correct answers for each course being tested. As of this writing, most schools are said to have made year-to-year progress; 15 percent have remained the same; and for 5 percent, performance has deteriorated. The schools that make the greatest gain are those whose performance levels were poorest at the outset. The data are used to design training programs for teachers, principals, and district coordinators and supervisors. Teacher training focuses on curricular matters and on the methodological and didactic aspects of learning (development of thinking skills, verbal interaction techniques, methods and models of content planning, thought processes related to that content, and strategies to link content and thinking skills). Supervisor training is geared to strengthening their pedagogical functions and taking them beyond purely administrative duties. An important dissemination element is the access parents are given to schools' annual results. Central and local authorities maintain that this strategy has fostered greater family participation in school life. The local media have engaged in a typically sensationalistic presentation of results, although the published information is gradually being treated more seriously from an educational perspective.

**International  
Tests**

Aguascalientes has not taken part in any comparative international tests independent of the national sample.

**Progress  
Observed**

Prominent among the advances made is the organizational coordination of the assessment system, inasmuch as it promotes the flow of information and joint work among schools, supervisors, and various management units of the Education Secretariat. The system has enjoyed a significant degree of external institutionalization, since teachers and trade unions support the initiative and collaborate in the assessment exercises. There has been substantial development of local technical capacities in assessing learning and analyzing data.

**Difficulties  
Observed**

The local technical teams have expressed some concern about the scale of the assessment initiative, given the limited human resources available. Some district supervisors have voiced concern about some parent groups, which use the test data to confront teaching staff rather than seek more effective means of collaboration.

**Outlook**

The budgetary constraints on expanding test coverage (in terms of both number of students and curricular areas assessed) has prompted a decision to focus technical efforts on ensuring that schools make the most use of the data available on sixth grade. There are hopes of using the test results to assess teachers, but the system lacks the mechanisms and professional incentives schemes that would be part of such assessments. In view of this proposal, sixth grade teachers question whether the tests are valid indicators of their professional performance and argue that the lower grades should also be subject to assessment. They also maintain that they should not be the only actors in the system to be held accountable for student academic performance. Beginning in the 2003–04 school year, census-based learning assessment was set to begin in the fourth grade of primary education.

---

## References

Asmar Amador, María Patricia, Mario Díaz Villa, and María Teresa Reyes Zambraño. 2002. *Elementos de Política para la Educación Superior Colombiana: Memorias de Gestión y Prospectiva*. Bogotá: MEN/ICFES.

Ferrer, Guillermo. 2004. *Las reformas curriculares de Perú, Colombia, Chile y Argentina: ¿Quién responde por los resultados?* Lima: GRADE.

Ferrer, Guillermo, with Gilbert A. Valverde and Juan Manuel Esquivel Alfaro. 1999. *Aspectos del currículum prescrito en América Latina: Revisión de tendencias contemporáneas en currículum, indicadores de logro, estándares y otros instrumentos*. Lima: PREAL, Working Group on Assessment and Standards.

Ferrer, Guillermo, and Patricia Arregui. 2003. *Las pruebas internacionales de aprendizaje en América Latina y su impacto en la calidad de la educación: Criterios para guiar futuras aplicaciones*. Document No. 26. Santiago: PREAL.

Ravela, Pedro. 2001a. “¿Cómo presentan sus resultados los sistemas nacionales de evaluación educativa en América Latina?” Document No. 22. Santiago: PREAL, Working Group on Assessment and Standards.

———, ed. 2001b. “Los próximos pasos: ¿Hacia dónde y cómo avanzar en la evaluación de aprendizajes en América Latina?” Document No. 20. Santiago: PREAL, Working Group on Assessment and Standards.

Rojas, Carlos, and Juan Manuel Esquivel. 1998. *Los sistemas de medición del logro académico en Latinoamérica*. LCSHD Paper Series No. 25. Washington, DC: World Bank, Latin America and the Caribbean Region, Department of Human Development.

Wolff, Laurence. 1998. *Educational Assessments in Latin America: Current Progress and Future Challenges*. Document No. 11. Santiago: PREAL.





---

# Interviews

Interviews were held with the following educational authorities, technical staff, and consultants during 2000–2004.

## National Cases

### ARGENTINA

Margarita Poggi

Jorge Fasce

Hilda Lanza

Silvia Montoya

Juan Cruz Perusia

Lucrecia Tulic

### BOLIVIA

Susana Barrera

Carlos Gutiérrez

Martha Méndez

Esther Balboa

Nicole Nucinkis

### BRAZIL

Alejandra Shulmeyer

INEP technical staff

### CHILE

Cristián Cox

Lorena Meckes

Silvia Elgueta

Patricia Aninat

Beatrice Ávalos

Leonor Cariola

Erika Himmel

### COLOMBIA

Carlos Pardo

Héctor Fernández

### COSTA RICA

Juan Manuel Esquivel

Sandra Arauz

### CUBA

Héctor Valdés

### DOMINICAN REPUBLIC

Leo Valeirón

### ECUADOR

Martha Grijalva

Gonzalo Barreno

### EL SALVADOR

Graciela de Salgado

Hilda Álvarez

### MEXICO

Felipe Martínez Rizo

Jesús Mejía

Víctor Velásquez

### PANAMA

Laura G. de Vergara

Ruth Nativí

### PARAGUAY

Martha Lafuente

### URUGUAY

Pedro Ravela

Beatriz Picaroni

Magela Figarola

### VENEZUELA

José Mirás

Julia Montoya

## Subnational Cases

MINAS GERAIS, BRAZIL

Cleuza Pereira

María Inés Barrozo

PARANÁ, BRAZIL

María Luiza Marques Dias

Marlene C. de Araujo

SÃO PAULO, BRAZIL

Cristina Amoroso

María Julia Filgueiras

Yara Espósito

BOGOTÁ, COLOMBIA

Eliana Iannini

José Medina

Álvaro Leuro

AGUASCALIENTES,

MEXICO

Javier Anaya

Josefina Mercado

Flabiano Jiménez



