

La evaluación de aprendizajes en América Latina

Felipe Martínez Rizo

No. 32

Cuadernos de Investigación



Este texto puede consultarse en: www.inee.edu.mx

NOVIEMBRE, 2008

CONTENIDO

Introducción	5
Los sistemas educativos de la región	7
Los sistemas de evaluación de los países latinoamericanos	9
Proyectos internacionales de evaluación en Latinoamérica	23
Conclusión	25
Referencias bibliográficas	27

INTRODUCCIÓN

Las evaluaciones en gran escala del aprendizaje de los alumnos comenzaron a utilizarse hace poco tiempo en diversos sistemas educativos. El caso más temprano fue el de los Estados Unidos, al que siguieron otros países del ámbito anglosajón. En general, el uso de tal tipo de evaluaciones se extendió a los países de América Latina hasta la última década del siglo XX. Algunos países comenzaron a aplicar pruebas estandarizadas en gran escala varias décadas antes, pero inicialmente para la selección de aspirantes a carreras universitarias de alta demanda; en lo relativo a la educación básica el uso de pruebas fue posterior y, en sus inicios, no sistemático.

En México comenzaron a utilizarse pruebas estandarizadas para la admisión de alumnos a algunas carreras en instituciones de educación superior como la UNAM, la UIA y el ITESM, desde la década de 1960; en el último caso utilizando pruebas en español elaboradas por la oficina del *College Board* de Puerto Rico. Las primeras evaluaciones en gran escala de educación básica se aplicaron desde la primera mitad de la década de 1970, pero fue hasta la de 1990 cuando el uso de ese tipo de instrumentos cobró importancia creciente para la educación básica. Además, las pruebas eran de baja calidad y sus resultados no se daban a conocer.

En Colombia, desde 1968 se estableció el Instituto Colombiano de Fomento de la Educación Superior (ICFES), para hacer pruebas de ingreso a las universidades. Sólo más de veinte años después se le encomendó la aplicación de pruebas de rendimiento para educación básica. En Costa Rica se comenzaron a aplicar pruebas a alumnos de educación básica desde los años setenta, a cargo de la Universidad de Costa Rica. En México comenzó a hacerlo en la misma década la Secretaría de Educación Pública. En Cuba,

desde 1972, con el programa *Perfeccionamiento del Sistema Nacional de Educación*, se pusieron en práctica mecanismos para monitorear y evaluar sus resultados y se fortaleció el sistema de inspección, que tenía ya larga tradición (Torres, 2008).

En todos los casos anteriores no se contaba realmente con un sistema de evaluación. Chile fue el país de la región que desarrolló más tempranamente un verdadero sistema; en 1968 se habían comenzado a utilizar pruebas en gran escala, y en 1978 el gobierno encomendó a la Pontificia Universidad Católica que desarrollara un sistema de evaluación. De manera más amplia fue hasta fines de los años ochenta y, sobre todo, durante los noventa, cuando los países de la región, de manera acelerada, comenzaron a implementar sistemas de evaluación de concepción moderna y alcance nacional. La tendencia ha continuado durante la última década, al grado que a fines de 2008 la mayor parte de los países de la región cuentan con sistemas nacionales de evaluación, si bien con grados de consolidación diversos.

Hay varias fuentes de información sobre las evaluaciones de aprendizaje que se aplican en los sistemas educativos de los países de América Latina y el Caribe. La información que contienen no tiene el mismo grado de actualización ni de confiabilidad, por lo que no siempre es consistente. En este trabajo se ha utilizado la información más reciente y completa; en muchos casos ha sido posible verificarla directamente con los responsables de cada país; por ello puede considerarse, en conjunto, más confiable y precisa que la de cada una de las fuentes utilizadas. En cada caso se citan las referencias consultadas.

Antes de describir los sistemas de evaluación de cada país, se presentan brevemente las características de los respectivos sistemas educativos. Poste-

riormente se destacan los rasgos coincidentes y discrepantes de los sistemas nacionales de evaluación, se presentan los esfuerzos regionales de evaluación,

y se sintetiza la experiencia de los países de la región en lo que se refiere a su participación en proyectos internacionales de evaluación.

1. LOS SISTEMAS EDUCATIVOS DE LA REGIÓN

Los sistemas educativos de América Latina y el Caribe tienen rasgos similares, seguramente debido a que comparten la influencia común de los sistemas de países europeos de los que la mayor parte de los países de la región fueron colonias desde principios del siglo XVI hasta la segunda década del XIX, en especial de España y Portugal.

El origen de los sistemas educativos modernos en los países de la región se remonta a los años que siguieron a su independencia política, en especial debido a legislaciones de orientación liberal, que en distintas fechas, a mediados del siglo XIX, sustituyeron formalmente el control de la Iglesia Católica sobre las escuelas por sistemas públicos a cargo del Estado. La precaria situación política y económica de la región, sin embargo, hizo que en casi todos los países el desarrollo efectivo de los sistemas educativos se diera hasta bien entrado el siglo XX.

Lo anterior explica que, en casi todos los países a los que se refiere este trabajo, la duración de la educación obligatoria comprendiera sólo la enseñanza primaria, en general de seis años, hasta las décadas de 1970 y 1980. Sólo posteriormente la obligatoriedad comenzó a extenderse hasta el final del primer ciclo de la secundaria, o inclusive hasta el final del segundo ciclo de esta etapa educativa.

Los sistemas educativos latinoamericanos comparten otros rasgos, herencia también de los sistemas ibéricos, como la práctica generalizada de hacer repetir grados escolares a muchos alumnos, por considerarse que no han alcanzado el nivel adecuado para pasar al grado siguiente. Esto, junto con la situación de pobreza de importantes sectores de la población, en una región que se distingue por altos niveles de desigualdad, ha traído consigo

la consecuencia de tasas reducidas de cobertura incluso en la educación primaria, que todavía no consiguen terminar todos los niños en la mayor parte de los países de la región.

En el contexto de las similitudes a las que se acaba de hacer referencia, los sistemas educativos latinoamericanos tienen también diferencias considerables, al grado de que un importante estudio reciente considera posible afirmar que no hay dos países, de los 19 más importantes de la región, que tengan igual estructura. Con información de la misma fuente (LLECE, 2008: 33) se presentan los datos siguientes.

La educación primaria es obligatoria en todos los países de la región y cada vez son más los que incluyen la secundaria inferior en esa condición. La obligatoriedad comienza a los cinco o seis años de edad en casi todas partes; México es una excepción, pues se comienza a los tres; en Guatemala y El Salvador, en cambio, comienza a los siete años, como ocurría en Brasil hasta 2006; ahora comienza también a los seis años.

En cuanto al fin de la obligatoriedad, ésta llega hasta los 12 años en Nicaragua y Panamá; termina a los 14 en Bolivia y República Dominicana; a los 15 en Costa Rica, Cuba, Ecuador, Honduras, México, Paraguay y Venezuela; a los 16 en El Salvador, Guatemala y Uruguay; a los 17 en Perú; y a los 18 años en Argentina, Brasil, Chile y Colombia.

En lo que se refiere a la forma en que se estructuran los niveles del sistema educativo, en algunos países de la región, los niveles de primaria y secundaria básica (1 y 2 según la Clasificación Internacional Normalizada de Educación de la UNESCO, CINE o ISCED) se integran en un mismo plantel, de manera que los alumnos asisten a un mismo estable-

cimiento durante ellos, lo que implica entre ocho y nueve años. Después los estudiantes pasan a la secundaria superior (CINE 3).

En otros países (Ecuador, Colombia, Cuba, Guatemala, México, Nicaragua, Panamá y Uruguay) la primaria y la secundaria básica constituyen dos niveles diferentes, con duración, en general, de seis y tres o cinco y cuatro años.

En la mayoría de los casos la educación formal comienza con el nivel de la educación preescolar o infantil (CINE 0), entre los tres y los cuatro años, y cubre al menos dos ciclos de educación preescolar. La asistencia a los primeros años de este nivel de enseñanza es optativa en todos los países salvo México que, como se ha dicho tiene cobertura obligatoria desde los tres años. Varios países más han establecido el último año de educación preescolar como obligatorio para el Estado, aunque no para las familias.

La tabla N° 1 sintetiza los principales rasgos de los sistemas educativos de los 19 países más importantes de la región.

Un dato más, fundamental para entender los sistemas educativos de los países latinoamericanos, es el que tiene que ver simplemente con su tamaño, en términos del número de alumnos que atiende el sistema, así como el número de escuelas y maestros que se requieren para ello. El número de alumnos depende a su vez, obviamente, del total de habitantes de cada país y de la estructura de edades de su población.

En los países de América Latina, la proporción de sus habitantes que tienen la edad en la que deben asistir a los niveles obligatorios de la escuela, es considerablemente más alta que en los países más desarrollados, como consecuencia del crecimiento demográfico que han experimentado durante la segunda mitad del siglo XX, cuyos efectos son todavía claramente perceptibles en la pirámide de edades.

Tabla N° 1. Estructura de los sistemas educativos de países de América Latina

País	Estructura del Sistema Educativo				Obligatoriedad		
	Preescolar	Primaria	Secundaria Básica	Secundaria Superior	Edad inicial	Edad final	Duración en años
	CINE 0	CINE 1	CINE 2	CINE 3			
Argentina	3	6	3	3	6	18	12
Brasil	2	9		3	6	18	12
Bolivia	2	8		4	6	14	8
Chile	3	8		4	5	18	13
Colombia	3	5	4	2	5	18	13
Costa Rica	1	9		2	5	15	10
Cuba	1	6	3	3	5	15	10
Ecuador	2	6	3	3	6	15	9
El Salvador	3	9		2	7	16	9
Guatemala	3	6	3	2	7	16	9
Honduras	3	3	6	2	6	15	9

Fuente: UNESCO-OREALC. 2008.

2. LOS SISTEMAS DE EVALUACIÓN DE LOS PAÍSES LATINOAMERICANOS

2.1 ARGENTINA

En 1993 el Ministerio de Educación del gobierno federal estableció un Sistema Nacional para la Evaluación de la Calidad Educativa, y desde esa fecha se aplican Operativos Nacionales de Evaluación (ONEs) cuyo alcance y cobertura se han ido ampliando. Inicialmente se aplicaban a alumnos del último grado de educación primaria y secundaria, e incluían sólo Matemáticas y Español; después se han extendido a otros grados y han incluido las Ciencias Naturales y las Ciencias Sociales. Se aplican también cuestionarios a directores, maestros, alumnos y padres de familia, y se hacen análisis sobre la influencia de los factores asociados con el rendimiento escolar, utilizando técnicas estadísticas sofisticadas.

Las pruebas para evaluar la primaria (Educación General Básica) se aplicaron hasta 1999, cuando tuvo lugar el 7° Operativo Nacional de Evaluación Muestral de EGB. En 1997 se aplicó por primera vez el ONE de Finalización del Nivel Secundario, que se

repitió en 1998, 1999 y 2000, en las cuatro ocasiones con carácter censal. En ese año se aprobó la creación de un órgano descentralizado, el Instituto de Calidad Educativa, con cierta autonomía, para que se encargara de las evaluaciones. La grave crisis por la que atravesó el país impidió que dicho organismo se pusiera en marcha efectivamente y el ONE previsto para 2001 no pudo llevarse a cabo.

En 2002 y 2003 los ONE continuaron, y a partir de la última fecha se decidió hacerlos cada dos años, por lo que los más recientes tuvieron lugar en 2005 y 2007, bajo la responsabilidad de la Dirección Nacional de Información y Evaluación de la Calidad Educativa (DNIECE) del Ministerio de Educación.

Muchas de las 24 jurisdicciones del país aplican además evaluaciones propias, de diferente concepción y cobertura. En las aplicaciones de 2007 la muestra de primarias comprendió cuatro mil escuelas y 75 mil alumnos, y la de secundarias dos mil escuelas y 45 mil alumnos.

Los Operativos Nacionales realizados de 1993 a 2007 se resumen en la tabla siguiente.

Tabla N° 2. Operativos Nacionales de Evaluación de Argentina

Año	Grados	Áreas	Cobertura
1993	7°, 5M, 6M	L, M	Muestra nacional
1994	7°, 5M, 6M	L, M	Muestra nacional
1995	3°, 9°, 5M	L, M	Muestra nacional
1995	6M, 7	L, M, CN, CS	Muestra nacional
1996	3°, 9°	L, M	Muestra nacional
1997	5M, 6M	L, M	Censo nacional
1998	6°, 7°	L, M, CN, CS	Muestra nacional
1999	3°, 7°, 9°	L, M	Muestra nacional
	6°	L, M, CN, CS	Muestra nacional
	5M, 6M	L, M, CN, CS	Censo nacional
2000	3°	L, M, CN, CS	Muestra nacional
2000	6°, 5M, 6M	L, M, CN, CS	Censo nacional
2002	3°, 6°, 9°, 5M	L, M	Muestra nacional y jurisdiccional
2003	3°, 6°, 9°, 12°	L, M	Muestra nacional y jurisdiccional
2005	3°, 6°, 9°, 12°	L, M	Muestra nacional y jurisdiccional
2005	3°, 6°, 9°, 12°	CN, CS	Piloto
2007	3°, 6°, 9°, 12°	L, M, CN, CS	Muestra nacional y jurisdiccional

3°, 6°, 7° y 9° = grados de Primaria; 5M y 6M = años de Enseñanza Media.

L = Lenguaje; M = Matemática; CN = Ciencias Naturales; CS = Ciencias Sociales.

Fuentes: Ferrer, 2006: 55-56; González Jiménez, 2008 1: 54.

Hasta 2002 se utilizaron pruebas convencionales, con enfoque normativo y preguntas de opción múltiple. En 2003 se equipararon las pruebas de Lengua y Matemáticas Aplicadas de 1996-2003 en 6° grado de EGB y al final del nivel secundario.

Desde 2005 se emprendió un proceso de mejora técnica de las pruebas. Se definieron niveles de logro en relación con los Contenidos Curriculares Comunes, que cada autoridad jurisdiccional debe precisar para sus alumnos. De esta manera los resultados de las pruebas pueden presentarse en la forma de porcentajes de alumnos que alcanzan uno u otro nivel. Las aplicaciones se controlan y se aplican cuestionarios de contexto para alumnos, docentes y directores de escuela.

Argentina participó en el proyecto TIMSS de la IEA, aplicando en el año 2000 el llamado TIMSS-R a una muestra nacional. En 2001 se aplicaron las pruebas del proyecto PIRLS a una muestra de 150 escuelas de 4° de EGB. Ese mismo año el país participó en el estudio sobre educación cívica de la IEA (CivEd), con alumnos de 9° de EGB y último grado de

educación secundaria. En los tres casos anteriores, al menos hasta 2008 los resultados no habían sido procesados.

Argentina comenzó a participar en el proyecto PISA de la OCDE desde su primera vuelta, con una aplicación en 2002 (PISA Plus); los resultados tampoco fueron procesados localmente, pero se difundieron en el informe internacional correspondiente. Aunque el país se propuso participar en la ronda de 2003 no consiguió hacerlo, pero sí en la de 2006. Una vez más los resultados del país sólo se difundieron en el informe internacional. Argentina participó también en los dos estudios del Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación, en 1997 y 2006 (Ferrer, 2006: 55-59; GIP, por aparecer; González Jiménez et al., 2008, Vol. 1).

2.2 BOLIVIA

En 1996 se estableció un Sistema de Medición de la Calidad (SIMECAL), que se concebía como un órgano semiautónomo, pero en la práctica funcionó

dentro del Ministerio de Educación. En 1997 el SIMECAL comenzó la aplicación de pruebas de rendimiento en los grados 3° y 6° de primaria, en las áreas de lenguaje y matemáticas, contemplando también el análisis de factores asociados.

En el año 2000 las pruebas se aplicaron en 1°, 3°, 6° y 8° de primaria, así como en 4° de secundaria, y con sus resultados se estableció una Línea de Base como referencia para evaluaciones posteriores. Los nuevos instrumentos adoptaron un enfoque criterial, y se definieron tres niveles de logro. Se hicieron pruebas no sólo en Español, sino también en Quechua y Aymara, las dos lenguas indígenas que habla gran parte de la población. En 1999 se hizo también una Evaluación del Sistema de Formación Docente del país.

El SIMECAL siguió operando hasta 2004; luego los cambios gubernamentales trajeron consigo el desmantelamiento de su estructura, y en 2008 no se han puesto en marcha mecanismos alternativos de evaluación.

En el plano internacional, Bolivia ha participado sólo en el Primer Estudio del LLECE (1997) y en el estudio ALL (*Adult Literacy and Lifeskills*, 2001) (Ferrer, 2006: 60-63).

2.3 BRASIL

La evaluación en gran escala de la educación básica comenzó en Brasil a fines de la década de 1980; de 1990 a 1995 se llevaron a cabo algunos operativos, con muestras de alumnos de los grados 1°, 3°, 5° y 7° e inició el *Sistema de Avaliação da Educação Básica* (SAEB). En la gestión del Presidente Cardozo se decidió fortalecer el SAEB, poniéndolo a cargo de un organismo creado décadas antes para hacer investigación educativa, que en 1997 adquirió el estatus de órgano descentralizado, aunque no autónomo (*autarquia*): el *Instituto Nacional de Estudos e Pesquisas Educacionais* (INEP). A diferencia de otros organismos de evaluación, el INEP se encarga tanto de la evaluación de la educación básica y media como de la que se refiere a educación superior. Por ello, además de las pruebas para educación básica del SAEB, el INEP aplica el *Examen Nacional do Ensino Médio* (ENEM), y exámenes de egreso para las principales carreras del primer nivel de la educación superior, el *Examen Nacional de Cursos*, conocido como el PROVAO.

Inicialmente, las pruebas del SAEB evaluaban Lectura y Matemáticas, y se aplicaban cada dos años a muestras nacionales de alumnos de los grados 4° y 8° de primaria y 2° y 3° de secundaria. Así ocurrió en

1993, 1995, 1997, 1999 y 2001. La aplicación de 1999 comprendió a más de 360 mil alumnos de 2 mil 100 municipios de los 26 estados y el Distrito Federal.

El ENEM, por su parte, se aplica en forma voluntaria al final de la educación media superior y es cada vez más utilizado como criterio para definir el ingreso a la universidad. En el año 2000 el PROVAO cubrió 18 carreras y fue aplicado a casi 198 mil egresados de instituciones de educación superior de todo el país, que ofrecieron en conjunto 2 mil 888 programas de esas carreras. Con un reconocimiento creciente por la elevada calidad técnica de su trabajo el INEP se encarga también del manejo del sistema nacional de indicadores educativos. (véase INEP, 2001)

En la década de 2000 el interés por la evaluación ha aumentado. Se siguen aplicando las pruebas del SAEB, que fueron rediseñadas en 2001, para permitir la definición de niveles de logro y van acompañadas por cuestionarios de contexto para alumnos, maestros y directores de escuela. Su carácter muestral permite un buen control de las aplicaciones, y el uso de técnicas de equiparación permite hacer comparaciones a lo largo del tiempo. A estas pruebas se han añadido otras de tipo censal, con las que se busca tener resultados de cada plantel escolar y sustentar acciones de mejora en ese mismo nivel.

Desde 2005 la *Prova Brasil* se aplica cada dos años a alumnos de 4° y 8° de primaria, en Portugués y Matemáticas. La aplicación cubre todas las escuelas públicas urbanas del país (censo), lo que permite dar resultados por escuela y municipio además de por estado federal y para todo el país; involucra a más de cinco millones de alumnos y no incluye cuestionarios de contexto. La *Provinha Brasil* tiene como propósito medir la alfabetización de los alumnos después del primer grado de la primaria, con carácter formativo; se aplica voluntariamente cada año en escuelas públicas, y los resultados se procesan en cada escuela.

Además de los operativos nacionales, los 27 estados federales del país pueden desarrollar sus propias evaluaciones y algunas tienen experiencias interesantes al respecto, como Sao Paulo, Minas Gerais y Paraná.

El interés por la evaluación en distintos momentos se refleja en la participación en algunos proyectos internacionales. Así, en 1991 Brasil participó en una evaluación de Matemáticas que llevó a cabo el *Educational Testing Service* de los Estados Unidos. Brasil ha participado también en los dos estudios del LLECE (1997 y 2006), así como en las tres aplicaciones del proyecto PISA (2000, 2003 y 2006) y en la del TIMSS de la IEA de 2003.

La importancia que el país atribuye a su participación en estos ejercicios internacionales puede apreciarse por el hecho de que las autoridades han adoptado una ambiciosa meta para el sistema educativo brasileño, al establecer que, en el año 2021, los alumnos brasileños de 15 años deberán alcanzar una media similar a la que obtengan en promedio los estudiantes de los países de la OCDE. Para dar seguimiento a los avances que se hagan en dirección de esta meta, las escalas de las pruebas PISA se han equiparado con las del SAEB. (Ferrer, 2006: 64-68; GIP, por aparecer)

2.4 CHILE

Como se ha dicho, Chile es el país que más tempranamente desarrolló un verdadero sistema nacional de evaluación para la educación básica. A la fecha se sigue considerando que el sistema chileno es particularmente fuerte, y se distingue porque tuvo tempranamente carácter censal.

Desde la reforma educativa de la década de 1960, comenzó a pensarse en el desarrollo de pruebas para evaluar la calidad de la educación chilena. Una primera prueba se aplicó a alumnos del 8° grado de educación básica en 1968. El interés por este tipo de evaluaciones se renovó en el marco de la municipalización de las escuelas y, en ese contexto, el paso a control privado de una cantidad importante de escuelas públicas, procesos emprendidos por el régimen militar a partir de 1978. En ese año, el Ministerio de Educación encomendó a la Universidad Católica de Chile el desarrollo de un sistema de evaluación del rendimiento de los alumnos del sistema educativo.

Como resultado de ello, desde 1982 comenzaron a aplicarse las Pruebas de Evaluación del Rendimiento (PER) a los alumnos de los grados 4° y 8° de la educación básica. En 1988 se estableció el Sistema de Medición de la Calidad de la Educación (SIMCE), que en 1995 fue asumido directamente por el Ministerio de Educación, a través de su Unidad de Currículo y Evaluación. En 2003 una Comisión de Expertos, nombrada directamente por el Ministro de Educación, hizo una evaluación general del SIMCE, con base en la cual se hicieron ajustes en particular a la forma en que se utilizan los resultados, pero se mantuvo en lo fundamental el diseño de las evaluaciones.

En 1982 y 1984 se aplicaron por primera vez pruebas en los grados 4° y 8° de la educación básica. A partir de 1988, ya establecido el SIMCE, las pruebas se han aplicado con intervalos de dos años,

con algunas excepciones, y alternando los grados: las pruebas de 4° grado se aplicaron en 1988, 1990, 1992, 1994, 1996, 1999, 2002. Para el 8° grado se aplicaron pruebas en 1989, 1991, 1994, 1995, 1997, 2000, 2004. En cada aplicación se evalúa a todos los alumnos en Lengua y Matemáticas, y a una muestra en Ciencias Naturales, Historia y Geografía, además de medir actitudes de los alumnos. En 1993 se aplicaron por primera vez pruebas a estudiantes del 2° grado de la educación media, lo que volvió a hacerse en 1994, 1998, 2001 y 2003. A partir de 2006, las pruebas del SIMCE se aplican cada año a los alumnos del 4° grado de educación básica, y en forma alternada un año a los del 8° grado básico y otro a los del 2° medio.

Las pruebas cubren las áreas de Lectura, Matemáticas y Comprensión del Medio Natural, y su referente son los Objetivos Fundamentales y Contenidos Mínimos Obligatorios del marco curricular nacional, a partir de los cuales se definen Mapas de Progreso.

Hasta 2005 los resultados se daban en la forma del puntaje promedio obtenido por cada alumno. A partir de 2006, comenzando con las pruebas de 4° grado, se han definido también niveles de logro (Inicial, Intermedio, Avanzado) y se informa también sobre la proporción de alumnos que alcanza cada nivel.

Hasta 1998 el SIMCE utilizó la Teoría Clásica de las Pruebas. A partir de entonces se usan modelos de Teoría de Respuesta al Ítem y se aplican métodos de equiparación. El sistema chileno es reconocido por su calidad, y ha conocido mejoras técnicas importantes.

Chile es el país de América Latina con mayor participación en proyectos internacionales de evaluación, desde el estudio de seis áreas de la IEA (1971); IALS (1998); TIMSS (1998, 2003); PIRLS y CivEd (2001); PISA (2002, Plus y 2006) y LLECE (1997, 2006) (Ferrer, 2006: 69-74; GIP, por aparecer; González Jiménez et al., 2008, Vol. 2).

2.5 COLOMBIA

En 1968 se creó el Instituto Colombiano de Fomento de la Educación Superior (ICFES) que desde entonces, entre otras actividades, desarrolla y aplica cada año pruebas de selección para el ingreso de los alumnos a las distintas universidades del país, lo que le ha dado gran credibilidad social y le ha permitido adquirir experiencia en el desarrollo y aplicación de pruebas de selección.

Por ello, cuando la preocupación por la calidad educativa llevó a importantes reformas a principios

de la última década del siglo XX y surgió el interés por la evaluación en gran escala, se encomendó también al ICFES el desarrollo de pruebas para la evaluación de alumnos de educación básica.

A partir de 1991 el ICFES comenzó a aplicar pruebas a muestras nacionales de alumnos de 3°, 5°, 7° y 9° de primaria, en Lengua y Matemáticas, y luego en Ciencias Naturales y Ciencias Sociales. En 1997 la Dirección de Calidad del Ministerio de Educación y el ICFES pusieron en marcha el Sistema Nacional para Evaluar la Calidad Educativa, conocido como SABER, que pretendía seguir una cohorte de alumnos en forma longitudinal con pruebas aplicadas anualmente, idea que se abandonó.

Las pruebas se siguieron aplicando a muestras de alumnos. En el ciclo 2002-2003 se aplicó el primer censo de pruebas de Lectura, Matemáticas, Ciencias Naturales y Competencias Ciudadanas. En el ciclo 2005-2006 tuvo lugar la segunda aplicación censal. Hasta 1998 se utilizó la Teoría Clásica de las Pruebas. A partir de entonces se usan modelos de Teoría de Respuesta al Ítem y se aplican métodos de equiparación. El sistema educativo de Bogotá cuenta también con un sistema de evaluación propio, y hay esfuerzos en el mismo sentido en otras regiones.

En 2008 el ICFES ha emprendido un importante trabajo de rediseño de sus pruebas, buscando mejorar su calidad técnica y medir de manera más precisa diferentes áreas curriculares, combinando aplicaciones censales y muestras, en un ciclo multianual de aplicaciones.

Colombia ha participado en los proyectos internacionales TIMSS (1995, 2000); PIRLS y CivEd (2001); PISA (2006) y LLECE (1997, 2006). (Ferrer, 2006: 75-79; GIP, por aparecer; González Jiménez et al., 2008, Vol. 3)

2.6 COSTA RICA

El Instituto de Investigación para el Mejoramiento de la Educación Costarricense (IIMEC), de la Universidad de Costa Rica, comenzó a llevar a cabo evalua-

ciones educativas en gran escala en 1986, mediante un acuerdo con el Ministerio de Educación Pública. En 1996 el Ministerio emprendió un proyecto para consolidar la evaluación educativa, que en 1997 llevó a la creación del Centro Nacional de Evaluación para la Educación (Ceneedu), que no pudo consolidarse, por lo que desde el año siguiente las evaluaciones son realizadas por un área del propio Ministerio.

La primera evaluación, en 1986, se hizo en los grados 3°, 6° y 9°, así como en el último de la enseñanza media, con carácter censal, en Lengua y Matemáticas. En 1987, 1990, 1996 y 1997 se llevaron a cabo otras aplicaciones que utilizaron muestras y cubrieron algunos de los mismos grados, en Lengua y Matemáticas, pero también en Ciencias Naturales y Ciencias Sociales. Se aplican también pruebas formativas en diversos grados y materias y pruebas sumativas en todas las materias a todos los alumnos que terminan el 9° grado, así como a los que terminan la enseñanza secundaria.

Por lo que se refiere a proyectos internacionales, Costa Rica ha participado en los dos estudios del LLECE (1997 y 2006) pero en el primer caso no llegó a la etapa de análisis y difusión de resultados (Ferrer, 2006: 80-83; Valeirón y Esquivel, 2006).

2.7 CUBA

El Sistema de Evaluación de la Calidad de la Educación (SECE) de Cuba comenzó a desarrollarse en 1996, al mismo tiempo que el primer estudio del Laboratorio Latinoamericano de Evaluación de la Calidad Educativa (LLECE).

Los Operativos Nacionales de Evaluación de la Calidad (ONEC) que lleva a cabo el SECE están a cargo del Instituto Central de Ciencias Pedagógicas (ICCP) del Ministerio de Educación (MINED). Hay además Grupos Provinciales y Municipales de Evaluación de la Calidad de la Educación. La tabla N° 3 muestra las principales características de las aplicaciones que se han realizado de 1996 a la fecha.

Tabla N° 3. Operativos Nacionales de Evaluación de la Calidad de Cuba

Año	Escuelas	Alumnos	Grados	Áreas
1996	99	3 726	3°	M, E
1996	100	---	3°	M, E
1997	100	3 766	3°	M, E
1999	200	8 000	3°, 6°	M, E
2001	280	12 500	4°, 6°, 7°, 9°	M, E
2002	6 115	34 615	7°, 9°, 12°	M, E
2003	876	19 498	3°, 5°, 6°, 7°, 8°, 9°	M, E
2004	1 349	93 032	2°, 4°, 6°, 7°, 8°, 9°	M, E
2004	384	24 435	3°, 6°, 10°, 11°, 12°	M, E, CN
2005	771	30 642	2°, 4°, 6°, 7°, 8°, 9°, 10°, 11°	M, E, CN, Ha
2006	840	73 392	7°, 8°, 9°, 10°, 11°	M, E, CN, Ha
2007	738	74 834	4°, 6°, 9°, 11°	M, E, CN, Ha

M = Matemáticas; E = Español; CN = Ciencias Naturales; Ha = Historia.

Se utilizan niveles de logro para reportar los resultados. Desde 1999 se aplican también cuestionarios de contexto a distintos actores; con datos derivados de ellos se elaboran informes para varias audiencias (docentes, directores de escuela, responsables provinciales, etc.), empleando un índice llamado socio-educativo y otro general, que comprenden dimensiones sobre la eficiencia del sistema, el desempeño profesional de los docentes y la inspección escolar, además del aprendizaje de los alumnos.

Una nueva versión del SECE, a partir de 2006, define seis principios básicos del Sistema: unidad de la precisión y la factibilidad de las metodologías utilizadas; participación masiva de los agentes educativos; carácter gradual de la evaluación; unidad de evaluación de variables producto y explicativas y unidad de la evaluación cuantitativa y cualitativa.

A partir de 2006 se evalúan también algunos grados de la Enseñanza Técnico-Profesional y de los programas en que se forman los maestros, y las evaluaciones incluyen la revisión de libretas de Matemáticas y Español de una submuestra de alumnos. Para 2008 se planeaba comenzar a evaluar la educación especial.

En el plano internacional Cuba ha participado únicamente en los dos estudios del LLECE (1997, 2006) (Ferrer, 2006: 84-86; González Jiménez et al., 2008, Vol. 4; Torres, 2008).

2.8 ECUADOR

De 1996 a 2000, la evaluación en gran escala de la educación básica se hizo mediante las pruebas del Sistema Nacional para la Medición del Rendimiento Académico (APRENDO) que llevaba a cabo un grupo externo al Ministerio de Educación, gracias a un financiamiento del Banco Mundial. Las pruebas se aplicaban a muestras nacionales de alumnos de los grados 3°, 7° y 10°, en las áreas de Lengua y Matemáticas.

A partir de 2001, al terminar el contrato con el Banco, la responsabilidad de las evaluaciones pasó al Ministerio de Educación, que siguió aplicando las pruebas en los mismos grados y áreas. En la última aplicación, en 2007, la muestra comprendió mil 817 escuelas y alrededor de 72 mil alumnos. Las áreas evaluadas fueron Lengua y Matemáticas en 3° y 6° de primaria y Ciencias (opcional) en 6° de primaria.

Las pruebas APRENDO usan la Teoría Clásica de las Pruebas y métodos de equiparación para poder tener resultados comparables en el tiempo. Las pruebas utilizadas en cada aplicación se hacen públicas después de ella. En 1998 y 2000, además de las pruebas se aplicaron los cuestionarios de contexto que acompañaron a las pruebas del primer estudio del LLECE; en 2007 se hizo lo mismo con los cuestionarios del Segundo Estudio.

A fines de 2006, y como parte de un Plan Decenal de Educación, se decidió sustituir dichas prue-

bas por un nuevo Sistema Nacional de Evaluación, cuyo desarrollo está a cargo de la Dirección General de Planeamiento del Ministerio de Educación.

Se planea que las nuevas evaluaciones se apliquen cada tres años con carácter censal en los grados 4°, 7°, 10° y 12° en Lectura y Matemáticas y, con carácter muestral, en Ciencias Naturales y Estudios Sociales de 7° y 10° grado. Se usarán pruebas matriciales, la Teoría de Respuesta al Ítem y cuestionarios de contexto. Se preveía que la primera aplicación tuviera lugar en 2008, con una participación de 23 mil 770 escuelas y 952 mil alumnos.

Ecuador ha participado en los dos estudios del LLECE (1997, 2006). (Ferrer, 2006: 91-93; GTEE, 2008, 4: 2-6)

2.9 EL SALVADOR

Las primeras evaluaciones en gran escala del sistema educativo salvadoreño, conocidas como pruebas SABE, tuvieron lugar de 1993 a 1996, en el marco de un proyecto para el fortalecimiento del rendimiento en educación básica. Cada año se aplicaron pruebas a una muestra nacional de algunos de los grados 3°, 4°, 5°, 6° y 9°, en Lengua, Matemáticas, Ciencias Naturales y Ciencias Sociales (véase tabla No. 4).

A partir de 2001, el Ministerio de Educación, a través de la Dirección Nacional de Monitoreo y Evaluación, puso en marcha el Sistema Nacional de Evaluación de los Aprendizajes (SINEA), que desde entonces lleva a cabo cada dos años pruebas de Logros de Aprendizaje (LA) que se aplican a alumnos de algunos grados de la educación básica. El SINEA asumió también la aplicación anual de las Pruebas

de Aprendizaje y Aptitudes para Egresados de Educación Media (PAES), que la Universidad Simeón Cañas aplicaba desde 1997 en forma obligatoria, y que eran un requisito para la obtención del título de bachiller. Los datos básicos de las aplicaciones más recientes son:

Las pruebas LA y PAES permiten definir niveles de logro. A partir de 2005, las PAES comenzaron a tener un peso de 20% en la calificación final de los futuros bachilleres; en 2007 su peso aumentó, pasando a ser de 25%.

El Salvador ha participado en los dos estudios del LLECE (1997, 2006). (Ferrer, 2006: 94-98; GTEE, 2008, 3: 2-6; Valeirón y Esquivel, 2006)

2.10 GUATEMALA

De 1992 a 1996 el Centro Nacional de Pruebas (CENPRE) del Ministerio de Educación, con apoyo de la Agencia Internacional para el Desarrollo de los Estados Unidos (USAID), en el marco del proyecto SABE, llevó a cabo las primeras evaluaciones a gran escala en educación básica.

En 1997, el Ministerio de Educación hizo un convenio con la Universidad del Valle de Guatemala, que asumió las tareas del CENPRE creando el Sistema Nacional para la Medición del Logro Académico (SINMELA), que luego se denominó Programa Nacional de Evaluación del Rendimiento Escolar (PRONERE), que formaba parte de una reforma educativa más amplia y contó con financiamiento del Banco Mundial. En 2004 el Ministerio solicitó al PRONERE emprender nuevas evaluaciones de algunos grados de primaria, así como de los alumnos que termina-

Tabla N° 4. Evaluaciones aplicadas en El Salvador, 2001-2007

Año	Grados	Áreas	Cobertura	Escuelas	Alumnos
2001	3°, 6°, 9°	L, M	Muestra	406	
2003	3°, 6°, 9°	L, M, CSMA, ESC	Muestra	471	42 146
2005	3°, 6°, 9°	L, M	Censo		349 483
2006	1°	L, M	Muestra	500	19 966
2007	PAES	L, M, CSMA, ESC	Censo		58 491

L = Lenguaje; M = Matemática; CSMA = Ciencia, Salud y Medio Ambiente; ESC = Estudios Sociales y Cívicos.

ban la educación secundaria; estas últimas se denominan pruebas de *Graduandos*. A partir de 2005 el Ministerio retomó estas actividades, estableciendo para ello el Sistema Nacional de Evaluación e Investigación Educativa (SINEIE), que está a cargo de la Dirección General de Evaluación, Investigación y Estándares Educativos (DIGEDUCA).

Desde la creación del SINEIE, cada año se han evaluado las competencias en Lectura y Matemáticas de alumnos de primaria (grados 3°, 6° y en una ocasión 1°). En general esas evaluaciones han sido muestrales. En forma simultánea se han aplicado las pruebas *Graduandos* a los alumnos que terminan el último grado de la enseñanza secundaria. Aunque no influyen en el promedio final, estas pruebas tienen carácter obligatorio, por lo que su aplicación es censal. La tabla N° 5 resume la información sobre las pruebas aplicadas en Guatemala.

Las pruebas más recientes tratan de evaluar competencias complejas y se desarrollan en español y en cuatro lenguas indígenas. Desde 2006 su diseño se hace con un enfoque criterial, con base en el cual se definen cuatro niveles de logro. Los resultados de las pruebas censales de *Graduandos* se dan por escuela; los planteles se ordenan según una estimación de los resultados de cada uno, te-

niendo en cuenta el nivel socioeconómico de los alumnos que atiende.

Las pruebas del primer estudio del LLECE se aplicaron en forma especial en Guatemala en 2002, con apoyo del Banco Mundial; el país participó en forma regular en el Segundo Estudio (2006) y ahora lo hace en el proyecto ICCS de la IEA. (Ferrer, 2006: 99-101; GTEE, 2008, 2: 2-4; Valeirón y Esquivel, 2006)

2.11 HONDURAS

De 1990 a 1994 tuvieron lugar los primeros intentos de llevar a cabo evaluaciones en gran escala, como parte del Programa de Mejora de la Educación Primaria, apoyado por la USAID. En 1995 se creó la Unidad de Medición de la Calidad de la Educación (UMCE), de la Universidad Pedagógica Nacional Francisco Morazán, que asumió la responsabilidad de las actividades de evaluación en el país, por encargo de la Secretaría de Educación Pública.

El primer levantamiento de datos tuvo lugar en 1997, cubriendo los grados 3° y 6° de primaria y las áreas de Español y Matemáticas. En 1998 se realizó el segundo levantamiento de datos, extendiéndose a los grados 2°, 3°, 4° y 6° de primaria, en

Tabla N° 5. Evaluaciones realizadas en Guatemala, 1992-2007

Año	Grados	Áreas	Cobertura	Responsable
1992 a 1996	3°	L, M	Muestra	CENPRE-SABE
1997-2001	3°, 6°	L, M	Muestra	PRONERE-UV
2004	1°, 3°	L, M	Muestra	PRONERE-UV
2004	Graduandos	L, M	Censo	PRONERE-UV
2005	3°, 6°	L, M	Muestra	SINEIE
2005	Graduandos	L, M	Censo	SINEIE
2006	1°, 6°	L, M	Muestra	SINEIE
2006	3°	L, M	Censo	SINEIE
2006	Graduandos	L, M	Censo	SINEIE
2007	Preescolar	L, M	Muestra	SINEIE
2007	3°, 6°	L, M	Muestra	SINEIE
2007	Graduandos	L, M	Censo	SINEIE

las mismas áreas curriculares. En 1999 se evaluaron los grados 2°, 3°, 4° y 5°; en 2000, 2002 y 2004, sólo 3° y 6°, incorporándose en la última ocasión el área de Ciencias Naturales.

Honduras ha participado en los dos estudios del LLECE (1997 y 2006). (Ferrer, 2006: 102-105; Valseirón y Esquivel, 2006)

2.12 MÉXICO

Desde principios de la década de 1970, la Secretaría de Educación Pública (SEP) comenzó a realizar actividades de evaluación en gran escala, que llevaron a la creación de un área especial para ello, que después se denominó Dirección General de Evaluación (DGE).

Las primeras pruebas de este tipo se aplicaron en 1972 para decidir la admisión de alumnos en secundaria. A fines de esa década se hicieron las primeras aplicaciones de pruebas a muestras de alumnos de primaria, con el proyecto llamado *Evaluación del rendimiento académico de los alumnos de 4° y 5° grados de educación primaria*.

Durante la década siguiente no hubo avances cualitativos, y se siguieron utilizando las pruebas que se habían comenzado a desarrollar previamente. A principios de los años noventa la experiencia relativa a evaluaciones en gran escala incluía los exámenes de ingreso a la educación normal, los de Diagnóstico para Alumnos de Nuevo Ingreso a Educación Secundaria (IDANIS) que se aplicaban en sólo tres entidades federativas, y el Estudio Integral de la Evaluación Preescolar, Primaria y Secundaria, aplicado en nueve estados.

A principios de la década de 1990 la evaluación en gran escala recibió un importante impulso, por la conjunción de varios elementos, y se concretó en varios proyectos:

- El de mayores dimensiones surgió en 1992, a partir de la firma del Acuerdo Nacional para la Modernización de la Educación Básica (AN-MEB), entre el gobierno federal, los gobernadores de los estados y el Sindicato Nacional de Trabajadores de la Educación.

De dicho acuerdo se derivaron medidas tan importantes como la descentralización del sistema educativo y el programa de estímulos económicos para docentes denominado Carrera Magisterial. Para asignar los estímulos se decidió tomar en cuenta, entre otros elementos, los resultados de los alumnos de los maestros que aspiraban a recibir apoyo del programa; ello trajo consigo la necesidad de aplicar cada

año pruebas a grandes números de alumnos de primaria y secundaria. La primera aplicación involucró a más de cuatro millones. Estas evaluaciones siguieron aplicándose hasta el año 2005, cuando el número de alumnos evaluados llegó a cerca de ocho millones.

- En 1991 el gobierno mexicano, con apoyo del Banco Mundial, promovió un programa educativo compensatorio en beneficio de cuatro de los estados más pobres, el *Programa para Abatir el Rezago Educativo* (PARE), cuya evaluación incluyó la aplicación de pruebas de aprendizaje. Con diferentes nombres, otros programas compensatorios siguieron, siempre con un componente de evaluación que incluía la aplicación de pruebas a los alumnos beneficiados. A partir de 1994 se buscó extender la aplicación de pruebas similares en todas las entidades, en forma permanente, mediante el proyecto denominado *Estudio de Evaluación de la Educación Primaria* (EVEP).
- En 1994 el ingreso de México a la OCDE fue un indicador del propósito de las autoridades de integrarse a la vida económica y política internacional, incluyendo el interés por participar en proyectos internacionales de evaluación educativa: en 1995 el TIMSS de la IEA, del que al final se decidió impedir que se publicaran los resultados del país, que no eran buenos; en 1996 el Laboratorio Latinoamericano de Evaluación de la Calidad Educativa de la OREALC; y, a partir de 1997, el proyecto PISA, de la OCDE, cuya primera aplicación tuvo lugar en el año 2000.
- A partir de la reforma de planes y programas de estudio que tuvo lugar a principios de la década, en 1996 se emprendió un trabajo de definición de estándares curriculares, junto con el desarrollo de instrumentos de evaluación del aprendizaje alcanzado por los alumnos en relación con esos estándares. Esos instrumentos se denominaron *Pruebas de Estándares Nacionales*, y se aplicaron por primera vez en 1998. Desde entonces, y hasta 2004, se aplicaron cada año a muestras nacionales de alumnos de primaria y, desde 2000, también de secundaria. Los grados en que se aplicaron pruebas cambiaron cada año, alternándose una vez el 2° y el 4° y otra el 3° y el 5°; el 6° grado se evaluó todos los años. Se evaluó Lectura y Matemáticas, y las pruebas eran acompañadas por cuestionarios para obtener información sobre el contexto de la enseñanza y del hogar.

- Se siguió aplicando el Instrumento de Diagnóstico para Aspirantes de Nuevo Ingreso a Secundaria (IDANIS) y se desarrollaron pruebas para premiar a alumnos destacados, en especial en Ciencias, en el marco de las llamadas Olimpiadas del Conocimiento.

Con la primera década del siglo XXI, y en relación con la regularización de la vida democrática del país que se manifestó en las elecciones de julio de 2000, se dieron nuevos pasos en lo relativo a la evaluación educativa en México, con la creación, en agosto de 2002, del Instituto Nacional para la Evaluación de la Educación (INEE), como instancia dotada de autonomía técnica, cuya misión se definió como la de hacer evaluaciones independientes del Sistema Educativo Nacional que, además de tener mejor calidad técnica, se distinguieran por una amplia difusión de sus resultados, en lo que constituyó un cambio fundamental en relación con la tradición de no difusión prevaleciente.

Inicialmente el INEE recibió de la DGE las bases de datos de las Pruebas de Estándares Nacionales aplicadas de 1998 a 2002, cuyos resultados no habían sido analizados ni, obviamente, difundidos, para que llevara a cabo dichos análisis y continuara la aplicación de tales instrumentos. Tras constatar la existencia de serias fallas técnicas de dichas pruebas, el INEE decidió emprender el desarrollo de nuevos instrumentos.

En 2004, el INEE decidió desarrollar instrumentos que cubrieran de manera más amplia las áreas del currículo nacional, mediante pruebas de diseño matricial que se aplicarían a muestras representativas de los alumnos de todos los estados federales en un ciclo de cuatro años según el cual un año se evaluaría a los alumnos de 3° de secundaria, otro a los de 6° de primaria, uno más a los de 3° de primaria, y otro a los de 3° de la educación preescolar.

En 2005 se aplicaron las primeras de estas pruebas, que se denominaron *Exámenes de la Calidad y el Logro Educativos* (Excale). Ese mismo año, con apoyo técnico del INEE, la SEP decidió emprender el desarrollo de otras pruebas nuevas, que se pretendía sustituyeran a las que se venían aplicando: las del Programa de Carrera Magisterial, las de ingreso a secundaria, las que se usaban para evaluar programas compensatorios y las de las Olimpiadas del Conocimiento. Estas pruebas se denominaron *Exámenes Nacionales del Logro Académico en Centros Escolares* (ENLACE), y se aplicaron por primera vez en 2006. La aplicación, sin embargo, no fue como se había planeado, ya que en lugar de utilizar un ciclo multianual similar al de las pruebas Excale, se decidió

aplicarlas anualmente a los cuatro últimos grados de primaria y el último de secundaria básica. Tampoco dejaron de aplicarse todas las otras pruebas que se pretendía sustituir con las nuevas.

Por lo que se refiere a proyectos internacionales de evaluación, y además de la frustrada experiencia del TIMSS, México ha participado en los tres ciclos de PISA (2000, 2003, 2006) y en los dos del LLECE (1997, 2006). (Ferrer, 2006: 106-110; GIP, por aparecer; González Jiménez et al., 2008, Vol. 5; Martínez Rizo, 2002)

2.13 NICARAGUA

Como otros países centroamericanos, Nicaragua participó en un proyecto promovido por la Organización de Estados Iberoamericanos (OEI) y la Coordinación Educativa y Cultural Centroamericana (CECC) para el Establecimiento de Estándares para la Educación Primaria en la región, que incluye un componente de evaluación.

Desde 1996 comenzaron a aplicarse pruebas de rendimiento por la Dirección de Evaluación del Ministerio de Educación y como parte del proyecto APRENDE. En 1999 se creó el Sistema Nacional de Evaluación (SNE) para la educación básica y media, con apoyo de USAID y UNESCO. Las pruebas se aplican a muestras nacionales, en 3° y 6° grado, y en las áreas de Lengua y Matemáticas. El país no ha participado en evaluaciones internacionales. (Ferrer, 2006: 111-112; Valeirón y Esquivel, 2006)

2.14 PANAMÁ

La historia de la evaluación en gran escala en Panamá se remonta a los años de 1981 a 1985, cuando se aplicaron pruebas a una muestra nacional de alumnos, en el marco de un proyecto promovido por la Organización de Estados Americanos (OEA). En 1988 y 1992 el Ministerio de Educación llevó a cabo otros operativos, y en 1995 lo hizo el Centro para el Estudio de la Calidad Educativa (CECE). Panamá participó también en el proyecto regional centroamericana de la OEI y la CECC antes mencionado.

A partir de 1999, el Ministerio de Educación puso en marcha un Sistema Nacional de Evaluación de la Calidad Educativa (SINECE), que ha llevado a cabo evaluaciones en ese año, en 2000 y en 2001. Sin embargo, en todos los casos se ha tratado de aplicaciones de tipo piloto, que no se han consolidado con una masa crítica de personal calificado. A partir de 2005 se ha manifestado un nuevo interés.

Panamá participó en el Segundo Estudio Regional del LLECE, y se prepara para hacerlo en la aplicación de 2009 de PISA. (Ferrer, 2006: 113-116; Valeirón y Esquivel, 2006)

2.15 PARAGUAY

En 1995, el Ministerio de Educación y Cultura creó el Sistema Nacional de Evaluación del Proceso Educativo (SNEPE) que aplica pruebas de rendimiento, como sigue:

2.16 PERÚ

En 1996, dentro del Ministerio de Educación, se creó la Unidad de Medición de la Calidad (UMC), que desde entonces ha realizado seis operativos de evaluación. Los cuatro primeros, de 1996 a 2004, tuvieron carácter muestral, mientras que los dos últimos fueron censales (Evaluación Censal de Estudiantes, ECE), si bien en el primero de ellos, el de 2006, sólo cubrió 55% de las escuelas y 44% de los alumnos. En 2006 se evaluó a los alumnos de

Tabla N° 6. Evaluaciones realizadas en Paraguay, 1996-2006

Año	Grados	Áreas	Cobertura	Escuelas
1996	6°	Co, M	Muestra	
1997	3°	Co, M	Muestra	
1997	9°	C, R, M	Muestra	
1998	6°	Co, R, M, ES	Muestra	
1998	12°	LL, R, M	Muestra	
1999	9°	LL, R, M, ES	Muestra	
2000	12°	LL, R, M, ES	Muestra	
2001	3°	Co, R, M, VST	Escuela Viva	
2001	6°	Co, R, M, ES	Escuela Viva	1 150
2004	3°	Co, R, M, MNS		
2004	6°	Co, R, M, CN		
2006	12°	LL, M, CB	Muestra	378

Co = Comunicación; M = Matemática; C = Castellano; R = Redacción; ES = Estudios Sociales; LL = Lengua y Literatura; VST = Vida Social y Trabajo; MNS = Medio Natural y Salud; CN = Ciencias Naturales; CB = Ciencias Básicas.

Las pruebas permiten definir niveles de logro. En la tabla anterior conviene destacar las evaluaciones de escuelas en situación de alto riesgo que se hicieron en el año 2001, incluyendo a mil escuelas rurales y 150 urbanas, con un total de aproximadamente 48 mil alumnos, atendidas por el Programa Escuela Viva. Se ha hecho también una evaluación de la formación de los maestros.

Paraguay ha participado en los dos estudios del LLECE (1997, 2006). (Ferrer, 2006: 117-120; GTEE, 2008, 6: 7-10)

2° grado de primaria, en las áreas de Lectura y Matemáticas.

En 2007 se evaluó de nuevo censalmente a los alumnos de 2° grado (en esta ocasión con una cobertura mejor, de 91% de las escuelas y 80% de los estudiantes), y también a los alumnos de 4° grado de Educación Intercultural Bilingüe, en Quechua y Aymara.

Los datos básicos de esas evaluaciones se resumen en la tabla N° 7:

Tabla N° 7. Evaluaciones realizadas en Perú, 1996-2007

Año	Grados	Áreas	Cobertura	Escuelas	Alumnos
1996	4°	L-C, M	Muestra	1 420	4 500
1998	4°, 6°, 10°, 11°	L-C, M, CA, PS	Muestra		68 000
2001	4°, 6°, 10°	L-C, M	Muestra		33 000
2004	2°, 6°, 9°, 11°	L-C, M	Muestra	1 479	56 000
2006	2°	L	Censo	15 600	300 135
2007	2°, 4°	L, M	Censo	28 248	674 220

L-C = Comprensión lectora o Comunicación; M = Matemáticas; CA = Ciencia y ambiente; PS = Personal-social; FC = Formación ciudadana.

En 2007 comenzó también una evaluación longitudinal que seguirá la cohorte que entró a 1° de primaria en ese momento hasta que termine el 6° grado, en 2012, también en las áreas de Comprensión Lectora y Matemáticas.

Perú ha participado en los dos estudios del LLECE (1997, 2006), así como en la aplicación especial que siguió a la primera ronda de PISA (PISA Plus, 2002) (Ferrer, 2006: 121-123; GTEE, 2008, 5: 6-9)

2.17 REPÚBLICA DOMINICANA

El Sistema de Evaluación comenzó en 1991, con un subsistema de pruebas nacionales que se aplicaron por primera vez ese año a los alumnos de 4° grado de la educación básica; en 1992 se aplicaron también pruebas en el 8° grado, así como en el Tercer Ciclo de Educación de Adultos; en 1993 se añadieron pruebas para el último grado del bachillerato (12°). Desde sus inicios las pruebas se aplican censalmente en los grados referidos.

A partir de 1994 las pruebas se aplican cada año a todos los alumnos del último grado de la enseñanza elemental (8°) y el último de la educación media (12°). Desde 2005 se estableció una fecha oficial como día de la evaluación, en el que se hace un reconocimiento a las mejores escuelas del país y de cada distrito: el tercer miércoles de octubre.

El carácter censal de estas evaluaciones se asocia a su alto impacto, ya que los resultados obtenidos por cada alumno en la prueba que presenta, con un peso del 30%, se combinan con las calificaciones que recibe en su escuela, para decidir si puede pasar al siguiente nivel. El alto impacto de las pruebas hace que los resultados deban entregarse en plazos muy cortos, lo que ha traído consigo el desarrollo de la

capacidad técnica necesaria para ello, pero también otras implicaciones, como el empleo exclusivo de reactivos que se puedan calificar automáticamente, y el que no se utilicen modelos complejos de escalamiento y reporte de resultados.

La República Dominicana ha participado en los dos estudios del LLECE (1997 y 2006) y se prepara para hacerlo en PISA 2009. (Ferrer, 2006: 87-90; Va-leirón y Esquivel, 2006)

2.18 URUGUAY

El máximo órgano del sistema educativo es una entidad que no depende del ejecutivo, sino de un Consejo de personalidades nombradas por el Congreso: la Administración Nacional de Educación Pública (ANEP). Esta entidad estableció en 1995 la Unidad de Medición de Resultados Educativos (UMRE), para encargarse de las actividades de evaluación, a partir de investigaciones realizadas en los años previos por el sociólogo German Rama, quien en la fecha mencionada era el titular de la ANEP.

La UMRE, cuyo nivel técnico y eficiencia son reconocidos, realiza evaluaciones desde 1996. El grado evaluado de manera más regular es el 6° de primaria, buscando visualizar el aprendizaje acumulado por los alumnos al final del primer nivel de la educación básica.

Los alumnos de 6° de primaria fueron evaluados por primera vez, en forma censal, mediante pruebas de Lengua y Matemáticas aplicadas en cada centro escolar por los docentes. Evaluaciones censales similares de alumnos de 6° grado se hicieron en 1999, 2001 y 2005. En 2008 se añadió el área de Ciencias Naturales.

En forma menos regular se han evaluado otros grados. En 1998 se evaluó a los alumnos de 3° de primaria, en Lengua, Matemáticas, Ciencias Naturales y Ciencias Sociales. En 1999 se evaluó a los estudiantes de 3° del Ciclo Básico de Educación Media (9° grado), también en Lengua, Matemáticas, Ciencias Naturales y Ciencias Sociales.

En 2001 se llevaron a cabo evaluaciones de los niveles de competencias lingüísticas y matemáticas al inicio de la primaria, mediante pruebas aplicadas por los maestros a alumnos de educación inicial, así como de 1° y 2° de primaria. En 2007 se repitió este tipo de evaluación, pero sólo con alumnos de 1° de primaria. Se evaluó también en forma masiva, pero con carácter voluntario, a los alumnos de 6° grado de primaria, sin muestra controlada. El censo de 1° de primaria comprendió unos 53 mil alumnos, y la aplicación de 6° grado aproximadamente 50 mil.

Las pruebas incluyen preguntas cerradas y abiertas. Lo que distingue especialmente el trabajo de la UMRE es su continuado esfuerzo por llevar los resultados a los maestros de todas las escuelas del país, en el marco de importantes esfuerzos de actualiza-

ción, que utilizan los resultados de las evaluaciones y los materiales de las pruebas como insumos para promover la mejora de las prácticas docentes.

Uruguay ha participado en los dos estudios del LLECE (1997 y 2006), así como en dos ciclos de PISA (2003 Y 2006). (Ferrer, 2006: 124-128; GTEE, 2008, 1; GIP, por aparecer; González Jiménez et al., 2008, Vol. 6)

2.19 VENEZUELA

A fines de la década de 1990, en el Ministerio de Educación se puso en marcha un Sistema Nacional de Evaluación del Aprendizaje (SINEA), en cuyo marco en 1998 se aplicó una prueba de Lengua y Matemáticas a una muestra nacional de alumnos del 6° grado de la educación básica. Los instrumentos se desarrollaron con apoyo de la Universidad Central de Venezuela.

Esta experiencia no tuvo continuidad, y hasta 2008 no se habían llevado a cabo nuevas evaluaciones. Venezuela participó en el primer estudio del LLECE (1997), pero no en el segundo. (Ferrer, 2006: 129-131)

3. PROYECTOS INTERNACIONALES DE EVALUACIÓN EN LATINOAMÉRICA

Como primer trabajo de este tipo puede mencionarse el proyecto Estudios Conjuntos sobre la Integración Económica Latinoamericana (ECIEL), coordinado por Claudio de Moura Castro, que en los años setenta analizó el nivel de preparación de la fuerza de trabajo de varios países de América Latina, para lo cual utilizó pruebas de la IEA. (Véase de Moura Castro, 1980; Muñoz Izquierdo 1976; Wolff, 1998).

En Centroamérica fue importante el proyecto para establecer estándares educativos de la Coordinación Educativa y Cultural Centroamericana, financiado por la Organización de Estados Americanos y la Agencia Española de Cooperación Internacional. De 1991 a 1996, la Agencia para el Desarrollo Internacional de los Estados Unidos (USAID) desarrolló un proyecto denominado *Strengthening of Achievement in Basic Education* (SABE) que tuvo importante influencia en los sistemas educativos de los países centroamericanos.

De mediados de la década de 1990 en adelante, la creciente conciencia de la importancia de la evaluación educativa en América Latina ha hecho que en el subcontinente se desarrollen iniciativas multinacionales en el campo, entre las cuales pueden mencionarse el Programa Iberoamericano de Evaluación de la Calidad Educativa (PIECE) de la Organización de Estados Iberoamericanos; el Proyecto de Establecimiento de Estándares para la Educación Primaria de la propia OEI y la Coordinación Educativa y Cultural Centroamericana; el Programa de Promoción de la Reforma Educativa en América Latina y el Caribe, PREAL; el Proyecto Regional de Indicadores Educativos, derivado de la II Cumbre de las Américas, coordinado por el Ministerio de Educación de Chile; y el Forum Hemisférico de Evaluación Educa-

tiva, derivado de la misma II Cumbre, y coordinado a su vez por el INEP de Brasil.

Entre esos trabajos multinacionales de evaluación en el subcontinente, destaca sin duda el del Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE) de la Oficina Regional de la UNESCO para América Latina y el Caribe (OREALC) que, junto con el SACMEQ africano, constituye el proyecto evaluativo internacional más importante desarrollado hasta la fecha, después de los de la IEA y del PISA de la OCDE.

En 1997 tuvo lugar la aplicación de las pruebas del Primer Estudio del LLECE, en Lengua y Matemáticas, acompañadas por cuestionarios de factores asociados, a muestras de alumnos de 3° y 4° grados de educación básica. Se contó con la participación de Argentina, Bolivia, Brasil, Chile, Colombia, Cuba, Dominicana, Honduras, México, Paraguay y Venezuela. En 2006 se llevó a cabo el Segundo Estudio, en el que participaron Argentina, Brasil, Chile, Colombia, Costa Rica, Cuba, Dominicana, Ecuador, El Salvador, Guatemala, México, Nicaragua, Panamá, Paraguay, Perú y Uruguay.

El aumento del interés en la evaluación se ha reflejado en la creciente participación en otros proyectos internacionales, en especial las evaluaciones PISA de la OCDE. En la primera aplicación en 2000, sólo participaron México, que ya era miembro de la OCDE, y Brasil. En la aplicación especial siguiente (PISA Plus) se añadieron Argentina, Chile y Perú. En 2003 esos tres países no participaron, pero sí lo hizo Uruguay, además de México y Brasil. En 2006 se integraron de nuevo Argentina y Chile, más Colombia. Para 2009 planean hacerlo varios países más.

Como se ha dicho ya, varios países de América Latina han participado en proyectos de evaluación

de la *International Association for the Assessment of Educational Achievement* (IEA). Chile lo hizo desde 1971 en el *Six Subjects Study*; en 1998 en el IALS; en el TIMSS en 1998 y 2003; y en el PIRLS y CivEd en 2001. Colombia participó en TIMSS (1995, 2000); PIRLS y CivEd (2001). Argentina participó en el TIMSS-R en 2000 y en el PIRLS y el CivEd en 2001. Brasil participó en TIMSS en 2003. En 2008 está en curso una aplicación especial del estudio sobre educación cívica de la IEA para siete países latinoamericanos.

Paralelamente se ha desarrollado la cooperación entre los países de la región. Para el Segundo Estudio del LLECE los responsables se reunieron regularmente de 2004 a 2008, y ahora lo hacen para preparar el Tercer Estudio. Los seis países participantes en PISA 2006, con España y Portugal, se han reunido regularmente también; preparan un informe conjunto que aparecerá en 2009, y apoyan a los países de la región que se han incorporado a PISA por primera vez.

4. CONCLUSIÓN

Esta revisión de las experiencias en evaluación en gran escala del aprendizaje en América Latina permite identificar aspectos comunes o singulares, y sacar algunas conclusiones.

4.1 ASPECTOS TÉCNICOS

En las primeras etapas las pruebas utilizadas en todos los países eran convencionales, en el sentido de que se construían con un enfoque normativo, estaban integradas exclusivamente por preguntas de opción múltiple y el escalamiento de las respuestas y el análisis de los resultados se basaban en la Teoría Clásica de las Pruebas; muchas veces los resultados se daban solamente en la forma de porcentajes de respuestas correctas.

En muchos países, pero con grados de avance diferentes, esos instrumentos han dejado el lugar a otros elaborados con otras características:

- Se adoptan las Teorías de Respuesta al Ítem, especialmente el modelo Rasch.
- Se busca un enfoque criterial, tratando de definir varios niveles de logro o desempeño, si bien no siempre parecen utilizarse metodologías rigurosas para definir esos niveles y los respectivos puntos de corte.
- Los resultados se dan tanto mediante puntajes obtenidos con escalas basadas en la TRI, como mediante porcentajes de alumnos en los diversos niveles de logro.
- En algunos casos se utilizan diseños matriciales para la construcción de las pruebas, y también preguntas de respuesta construida.
- Con la información de los cuestionarios de contexto que se aplican con las pruebas, los resultados se desglosan no sólo por región, sino tam-

bién por género, nivel socioeconómico y grupo étnico.

- Comienzan a hacerse análisis más complejos, usando por ejemplo modelos multinivel.

4.2 LA DIFUSIÓN DE LOS RESULTADOS

Sin contar las pruebas de ingreso a la universidad u otros programas educativos, hasta mediados de la década de 1990 los resultados de las pruebas de aprendizaje en gran escala que se aplicaban en los países de la región latinoamericana no se usaban, en general, para tomar decisiones que afectaran a personas o instituciones en lo individual.

Las pruebas no se utilizaban para decidir sobre la aprobación o reprobación de los alumnos, ni tampoco para asignar estímulos o tomar medidas correctivas que afectaran a maestros o escuelas individuales. Por ello las pruebas podían definirse como de bajo impacto. En algunos casos su impacto era de hecho nulo, por la ausencia de difusión de los resultados.

La excepción más destacada fue el SIMCE de Chile, que desde sus inicios se definió como de alto impacto: el diseño censal de las pruebas se hizo con el propósito de contribuir a la introducción de cambios mayores en el sistema educativo, con su municipalización y su relativa privatización. Los resultados del SIMCE se utilizan para decidir cuáles escuelas pueden recibir fondos públicos, en la forma de bonos individuales para sus alumnos.

La proliferación de las pruebas en años más recientes, en especial ya en el siglo XXI, ha ido acompañada por un creciente interés porque sus resultados sean utilizados para sustentar decisiones de las que se espera se deriven mejoras importantes de la calidad educativa.

Esta tendencia se relaciona con la noción de rendición de cuentas (*accountability*), y cobra sentido en el contexto de otras corrientes más amplias: por una parte, la búsqueda de mayor transparencia en el manejo de los asuntos públicos; con frecuencia también, por otra parte, aunque con matices importantes, cierta desconfianza respecto de la educación pública y, en general, respecto a la gestión pública de los servicios. A lo anterior debe añadirse la escasa cultura que hay en las sociedades latinoamericanas en cuanto a evaluación educativa, y no sólo entre el público en general, sino también entre maestros y autoridades educativas, e incluso entre investigadores y especialistas.

La combinación de los elementos anteriores está llevando, en varios casos, a que se esperen resultados casi milagrosos de mejora de las escuelas gracias a la aplicación de pruebas de aprendizaje en gran escala, sin tener en cuenta sus reales alcances y sus limitaciones.

Quienes desconocen la complejidad de los procesos educativos tienden a pensar que la difusión de resultados de las pruebas en la forma de ordenamientos simples de escuelas, con base en los puntajes obtenidos por sus alumnos, sin más consideraciones, permitirá una competencia de la que se derivarán casi automáticamente mejoras importantes.

Frente a estas posturas simplistas, es importante reiterar que las pruebas en gran escala nunca podrán sustituir el trabajo de los maestros, incluyendo su dimensión evaluativa. La evaluación más importante, una que incluya todos los aspectos del currículo y los niveles cognitivos más complejos, teniendo en cuenta las circunstancias de cada alumno, sólo puede hacerla un buen maestro, en el contexto del aula, con acercamientos mucho más ricos que los que pueden emplearse en gran escala. Las pruebas de este tipo deben ser vistas como un complemento del trabajo de los maestros, y no como un sustituto de él.

Es fundamental un cambio de perspectiva, que deje de ver a la evaluación como amenaza externa, para pasar a considerarla una excelente oportunidad de aprendizaje.

Desde esta distinta perspectiva, una tendencia como la de aplicar pruebas censales podría entenderse de otra manera: en vez del propósito de hacer ordenamientos simples, los resultados podrían servir para detectar oportunamente alumnos en riesgo y escuelas que necesitan apoyo especial, para brindarlo oportunamente a unos y otras, en lugar de propiciar competencias estériles de las que se derivan consecuencias perversas como la de enseñar para la prueba.

4.3 CONSIDERACIONES FINALES

Pese a un inicio tardío, y a una débil tradición psicométrica, los sistemas de evaluación de los países latinoamericanos están consolidándose aprovechando la experiencia internacional. Los sistemas de Chile, México, Brasil y Uruguay ofrecen ya información valiosa para sustentar decisiones. Hay procesos de mejora técnica en Argentina, Colombia, Ecuador, El Salvador y Guatemala. El sistema de Cuba, muy distinto del resto, muestra una efectividad sorprendente, según los resultados del LLECE. La creciente cooperación regional puede contribuir al fortalecimiento de los sistemas más débiles.

Dos riesgos están presentes. La falta de continuidad que traen los cambios políticos, cuyo efecto destructor se ha constatado, y el peligro de los efectos perversos para la calidad educativa que traen consigo prácticas derivadas de una comprensión inadecuada de los alcances y limitaciones de las evaluaciones en gran escala. La reciente tendencia a hacer evaluaciones censales puede ser positiva, pero también un avance en dirección negativa.

En contraste con lo que pasaba hace una década, cuando las evaluaciones eran pocas y sus resultados casi no se difundían, hoy las aplicaciones proliferan y comienzan a cansar a las escuelas; los resultados se debaten, a veces con escaso conocimiento de su complejidad. Para que las perspectivas favorables puedan concretarse será necesario que una visión objetiva de las posibilidades de la evaluación prevalezca sobre las visiones superficiales.

REFERENCIAS BIBLIOGRÁFICAS

- De MOURA CASTRO, CLAUDIO (1980). La educación en América Latina: una síntesis de los estudios comparativos sobre costos y eficiencia. *Revista Latinoamericana de Estudios Educativos*. Vol. X, N° 4, p. 121.
- DIR. GRAL. DE EVALUACIÓN (1997b). *Reseña histórica de la evaluación educativa que ha desarrollado la SEP 1970-1994*. Anexo DGE, 1997a. México, SEP.
- FERRER, GUILLERMO (2006). *Educational Assessment Systems in Latin America: Current Practice and Future Challenges*. Washington. Programa de Promoción de la Reforma Educativa en América Latina y el Caribe, PREAL.
- GRUPO IBEROAMERICANO DE PISA (Por aparecer). *Resultados de PISA 2006 en Iberoamérica*. Cap. 2. Educación y contextos de los países iberoamericanos en PISA. Anexo. PISA en relación con los programas de evaluación de cada país.
- GONZÁLEZ JIMÉNEZ, D. ASTRID et al. (2008). *Evaluaciones externas en América Latina. Caracterización de los procesos logísticos utilizados para la aplicación de pruebas censales y muestrales en Cuba, Chile, Argentina, México, Uruguay y Colombia*. 6 vols. Cali-Bogotá. Universidad de San Buenaventura-ICFES.
- GRUPO DETRABAJO SOBRE ESTÁNDARES Y EVALUACIÓN (2007). Evaluaciones nacionales en el 2007. *Observatorio Regional de Políticas de Evaluación Educativa*. Boletín N° 1 (Diciembre). PREAL. Págs. 6-7.
- GTEE (2007). México. *Observatorio Regional de Políticas de Evaluación Educativa*. Boletín N° 1 (Diciembre). PREAL. Págs. 7-12.
- GTEE (2008). Guatemala. *Observatorio Regional de Políticas de Evaluación Educativa*. Boletín N° 2 (Febrero). PREAL. Págs. 2-4.
- GTEE (2008). El Salvador. *Observatorio Regional de Políticas de Evaluación Educativa*. Boletín N° 3 (Abril). PREAL. Págs. 2-6.
- GTEE (2008). Ecuador. *Observatorio Regional de Políticas de Evaluación Educativa*. Boletín N° 4 (Mayo). PREAL. Págs. 2-6.
- GTEE (2008). Perú. *Observatorio Regional de Políticas de Evaluación Educativa*. Boletín N° 5 (Junio). PREAL. Págs. 6-9.
- GTEE (2008). Paraguay. *Observatorio Regional de Políticas de Evaluación Educativa*. Boletín N° 6 (Agosto). PREAL. Págs. 7-10.
- HORN, R., L. WOLFF y E. VÉLEZ (1991). *Developing educational assessment systems in Latin America: A review of issues and recent experience*. Washington. World Bank.
- LABORATORIO LATINOAMERICANO DE EVALUACIÓN DE LA CALIDAD DE LA EDUCACIÓN (2008). *Los aprendizajes de los estudiantes de América Latina y el Caribe. Primer Reporte. Segundo Estudio Regional Comparativo y Explicativo*. Santiago de Chile. Oficina Regional de Educación para América Latina y el Caribe de la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura, OREALC-UNESCO.
- LABORATORIO LATINOAMERICANO DE EVALUACIÓN DE LA CALIDAD DE LA EDUCACIÓN (2000). *Primer estudio internacional comparativo sobre lenguaje, matemática y factores asociados, para alumnos del tercer y cuarto grado de la educación básica. Segundo informe*. Santiago de Chile. OREALC.
- MARTÍNEZ RIZO, FELIPE (1996). *La calidad de la educación en Aguascalientes. Diseño de un sistema de monitoreo*. Aguascalientes. UAA-IEA.
- McMEEKIN, R. W. (1998). *Estadísticas educativas en América Latina y el Caribe*. Informe Técnico.

- Washington. Banco Interamericano de Desarrollo.
- TIANA FERRER, ALEJANDRO. (2000). *Cooperación internacional en evaluación de la educación en América Latina y el Caribe. Análisis de la situación y propuestas de actuación*. Washington. Banco Interamericano de Desarrollo.
- TORRES FERNÁNDEZ, PAUL (2008). *Evaluación educativa: los estudios internacionales y su influencia en Cuba*. Trabajo presentado en las IV Jornadas de cooperación educativa iberoamericana sobre evaluación. Antigua, Guatemala.
- UNESCO (2008). *Education for All. Global Monitoring Report 2008*. Paris. UNESCO. National learning assessments by región and country. Pp. 208-220. Latin America and the Caribbean pp. 214-215.
- VALEIRÓN, JULIO y J. M. ESQUIVEL (2006). *Situación de la Evaluación y la Investigación Educativa en los Países Centroamericanos y la República Dominicana*. San José, Costa Rica. Coordinación Educativa y Cultural Centroamericana y Agencia Española de Cooperación Internacional.
- WOLFF, LAWRENCE (2004). Educational Assessments in Latin America: The State of the Art. *Applied Psychology: An International Review*. Vol. 53 (2), 192-214.
- WOLFF, LAWRENCE (1998). Educational assessments in Latin America: Current progress and future challenges. Washington. PREAL.