



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

N° 38

Los costos de las evaluaciones de aprendizaje en América Latina

Laurence Wolff*

Julio 2007

* Laurence Wolff se ha desempeñado desde 1997 como consultor del Banco Interamericano de Desarrollo, entre otras agencias y organizaciones. Con anterioridad, trabajó durante 22 años con el Banco Mundial. De nacionalidad norteamericana, tiene el grado de Dr. en Planificación Educativa del Harvard Graduate School of Education.

Las opiniones vertidas en este trabajo son responsabilidad del autor y no comprometen a PREAL ni a las instituciones que lo patrocinan.

© 2007 Programa de Promoción de la Reforma Educativa en América Latina y el Caribe (PREAL)

Todos los derechos reservados

Si desea solicitar copias adicionales, contacte al PREAL y al Inter-American Dialogue.
Este estudio puede ser descargado de: www.preal.org/publicacion.asp

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

Para citar: Wolff, Laurence, Los Costos de las Evaluaciones de Aprendizaje en América Latina
(Washington, DC: PREAL, 2007).

ISBN: 0-9772271-8-9
Primera Edición
Publicado en Chile

Diseño e Impresión:
Editorial San Marino
esm.cl

ÍNDICE

RECONOCIMIENTOS	5
ACRÓNIMOS Y ABREVIATURAS	6
RESUMEN EJECUTIVO	7
I. ¿POR QUÉ ESTIMAR LOS COSTOS DE LAS EVALUACIONES?	8
• EVALUAR ES FUNDAMENTAL	8
• TRES TIPOS BÁSICOS DE EVALUACIONES.....	9
• SÍNTESIS DEL ESTUDIO	10
II. LOS COSTOS DE LAS EVALUACIONES EN AMÉRICA LATINA	12
• PRUEBAS CENSALES	12
• PRUEBAS MUESTRALES	15
• PARTICIPACIÓN EN EVALUACIONES INTERNACIONALES	22
III. CONCLUSIONES SOBRE LOS COSTOS DE LAS EVALUACIONES EN AMÉRICA LATINA	25
• EL COSTO DE LAS EVALUACIONES COMO PORCENTAJE DE LOS COSTOS DE LA EDUCACIÓN	25
• VARIABILIDAD DE LOS COSTOS DE LAS EVALUACIONES	28
IV. COSTOS Y TOMA DE DECISIONES SOBRE LAS EVALUACIONES	31
• LA DIFICULTAD DE MEDIR RESULTADOS EN EDUCACIÓN	31
• EQUILIBRANDO LO DESEABLE Y LO POSIBLE (<i>TRADE-OFFS</i>) EN TÉRMINOS DE METAS, COMPLEJIDAD, CAPACIDAD Y COSTOS DE LAS EVALUACIONES	31
• ÉNFATIZANDO LA DIFUSIÓN DE LOS RESULTADOS	32
• EL IMPACTO POTENCIAL DE LAS PRUEBAS DE ALTAS CONSECUENCIAS	33
• EL VALOR DE PARTICIPAR EN PROGRAMAS INTERNACIONALES DE EVALUACIÓN	34
• NECESIDAD DE IDENTIFICAR “QUÉ FUNCIONA” EN LA EDUCACIÓN	34
BIBLIOGRAFÍA	35
• ANEXO: RESUMEN DE SISTEMAS NACIONALES DE EVALUACIONES DE APRENDIZAJES	37

Lista de tablas

1. Objetivos posibles de las evaluaciones realizadas en América Latina
2. Chile - Pruebas SIMCE
3. Colombia - Evaluación nacional en educación básica y examen de ingreso a la educación superior
4. Colombia - Prueba muestral de educación básica
5. Uruguay - Prueba muestral de educación primaria y media
6. Perú - Prueba muestral de educación primaria y secundaria
7. Honduras - Evaluación muestral
8. Uruguay - Participación en PISA
9. Perú - Participación en PISA
10. Colombia - Costos de participación en PIRLS
11. Costos estimados de la prueba LLECE de UNESCO 1997-99
12. Resumen de los costos de las evaluaciones en América Latina
13. Costo estimado de intervenciones potenciales en educación primaria
14. Costos de evaluación por área de estudio y grado evaluados

Lista de gráficos

1. Costos de evaluación como porcentaje del presupuesto total del nivel educativo correspondiente
2. Costos totales y costos ajustados
3. Costos totales y costos ajustados por alumno evaluado

Reconocimientos

El autor desea agradecer a los funcionarios e investigadores Lorena Meckes (Chile), Daniel Bogoya y Álvaro Díaz (Colombia), Renán Rápalo (Honduras), Liliana Miranda (Perú), y Pedro Ravela y Beatriz Picaroni (Uruguay) por su gentil aporte de información para el presente estudio y por comentar los borradores; a Lucrecia Palacios y Julien Hautier por su asistencia en el análisis de datos; a Jeff Puryear, Hans Wagemaker y Andreas Schleicher por sus comentarios a versiones preliminares; a Kristin Saucier por darle forma final al documento. El autor está sumamente agradecido a Patricia Arregui por su apoyo continuo al inicio y durante la preparación de este informe. Cualquier error o inconsistencia en los datos son responsabilidad del autor.

Acrónimos y Abreviaturas

ANEP	Administración Nacional de la Educación Pública - Uruguay
PBI	Producto Bruto Interno
ICFES	Instituto Colombiano para el Fomento de la Educación Superior - Colombia
BID	Banco Interamericano de Desarrollo
IEA	International Association for the Evaluation of Educational Achievement (Asociación para la Evaluación de los Logros Educativos)
LLECE	Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación
PIRLS	Progress in Reading and Literacy Study
PISA	Programme for International Student Assessment
PPA	Paridad del Poder Adquisitivo
SABER	Sistema Nacional de Evaluación de la Calidad de la Educación - Colombia
SIMCE	Sistema de Medición de la Calidad de la Educación - Chile
TIMSS	Trends in International Mathematics and Science Study
UMC	Unidad de Medición de la Calidad Educativa - Perú
UMCE	Unidad de Medición de la Calidad Educativa - Honduras
UNESCO	United Nations Educational, Scientific, and Cultural Organization

RESUMEN EJECUTIVO

La realización de evaluaciones, al menos a un nivel muestral, se está convirtiendo rápidamente en un elemento fundamental de los sistemas educativos modernos, dado que representa una parte necesaria del proceso de diseño, implementación y evaluación de programas para mejorar la calidad de la educación y, además, porque países de muchos lugares del mundo están realizando crecientemente evaluaciones censales, usualmente en determinados grados escolares.

Hasta el momento, se ha publicado muy poca información respecto al tema de los costos de las evaluaciones y ninguna sobre el caso particular de América Latina. El presente estudio tiene como propósito brindar estimados preliminares sobre este tema, basados en información provista por Chile, Colombia, Honduras, Perú y Uruguay.

Cada uno de estos países ha seguido su propio conjunto de criterios y definiciones, y no se ha realizado trabajo de campo para confirmar sus estimados. En particular, los costos indirectos de administración de una unidad de evaluación no han sido considerados. Los datos presentados deben ser, por lo tanto, considerados indicativos más que definitivos. Los datos disponibles corresponden a evaluaciones censales en Chile y Colombia, a evaluaciones muestrales en cuatro de los cinco países y a evaluaciones internacionales (muestrales) en Chile, Colombia y Perú.

Esta revisión condujo a los siguientes hallazgos:

- Los costos varían significativamente de país en país, incluso cuando se evalúa aproximadamente al mismo número de estudiantes. Esta variación responde a diversas razones, tales como las siguientes:
 - Algunos países evalúan varios grados y materias en un mismo momento o tienen una cobertura curricular más extensa y/o profunda, lo que incrementa los costos.
 - Algunos países recogen información sobre posibles factores determinantes del rendimiento, mientras que otros no lo hacen.
 - Los países con gran número de habitantes y amplia área territorial son más costosos de muestrear que países más compactos; es más costoso calificar respuestas a preguntas abiertas que a preguntas de opción múltiple.
 - Países con limitados recursos humanos pueden requerir de los servicios de consultores locales o internacionales, lo que resulta más caro.
 - Algunos países consideran deseable que sean los mismos profesores o funcionarios del sector quienes administran las pruebas, mientras que otros prefieren contratar examinadores externos.
 - Las escalas salariales pueden variar de país en país aun luego de considerar la paridad del poder adquisitivo (PPA).

Las comparaciones de costos –que no son el foco central de este estudio– deberían tomar en cuenta dichas diferencias¹.

- Los costos absolutos de realizar evaluaciones fluctúan entre \$ 111,000 (PPA) por una evaluación muestral en Uruguay hasta \$ 6.5 millones en Chile por una evaluación aplicada a todos los estudiantes de octavo grado². Comprensiblemente, las evaluaciones censales son más costosas que las muestrales. Los costos de participación en programas de evaluación internacionales tales como el Programa para la Evaluación Internacional de Estudiantes (PISA) y Estudio sobre Tendencias Internacionales

1 Se ha incluido unas tablas en el anexo para facilitar dichas comparaciones.

2 Todos los valores monetarios señalados en este estudio están expresados en su equivalencia en dólares americanos convertidos utilizando la paridad del poder adquisitivo (PPA).

en Matemáticas y Ciencias (TIMSS) están en un rango de \$ 311,000 a \$ 599,000 en 3 países.

- Los costos por alumno obviamente varían dependiendo del número de alumnos evaluados. Van desde \$2.50 por alumno en una evaluación de más de un millón de estudiantes en Colombia hasta alrededor de \$100 por alumno en Perú y Colombia por una evaluación muestral de 5,000 alumnos en una prueba internacional.

A pesar de estas diferencias, los costos de evaluaciones tal y como se realizan en la región (por ejemplo, pruebas muestrales o evaluaciones censales a determinados grados) no representan una carga pesada para los presupuestos educativos. En ninguno de los países estudiados las evaluaciones representaron más del 0.3% del presupuesto educativo del nivel evaluado (por ejemplo, primario o secundario).

Sobre la base de los datos provistos, así como de la revisión de literatura sobre el impacto de las evaluaciones en el rendimiento académico, se puede arribar a las siguientes conclusiones:

- Evaluar está entre las innovaciones menos costosas en la reforma educativa primaria, teniendo un costo mucho menor al del aumento de los salarios docentes, la reducción del tamaño de los salones de clase o la reforma de la formación docente.
- Los costos juegan un rol importante, pero no determinante, en la toma de decisiones sobre evaluación. Cada país tiene un conjunto de condiciones diferente, y los tomadores de decisiones y los técnicos necesitan tomar sus propias decisiones (*trade-offs*) sobre la amplitud y profundidad de las mediciones en función a sus propios objetivos y capacidades. Dadas las actuales capacidades, no es recomendable evaluar a todos los estudiantes de todos los grados, tal como está normado actualmente en Estados Unidos.
- Las únicas situaciones en las que las evaluaciones pueden ser consideradas como un mal uso de fondos públicos es cuando la calidad técnica de las pruebas es tan baja que no permite arribar a conclusiones válidas sobre el aprendizaje, o cuando la información obtenida con las pruebas no se disemina ni utiliza. Los tomadores de decisiones no deben subestimar la compleji-

dad y los retos técnicos que implica la medición del logro académico y deben realizar todos los esfuerzos posibles por asegurar alta calidad técnica. Asimismo, debe contarse con un plan de diseminación completamente financiado antes de comenzar cualquier programa de evaluación.

- Hay creciente evidencia internacional que sustenta que las pruebas de altas consecuencias, o los exámenes de salida basados en el currículo, diseñados para certificar que los estudiantes han completado exitosamente un nivel de educación (usualmente la secundaria), pueden elevar el nivel de logro académico. Los países podrían considerar esta opción como un medio para mejorar el rendimiento académico en la educación secundaria.
- La participación en pruebas internacionales no resulta costosa, y puede resultar muy rentable si se usan los resultados para reformar el currículo y los programas de formación docente.
- Se debe invertir más dinero en medir “aquello que funciona” en educación, incluyendo, pero no restringiéndose a los posibles impactos de las evaluaciones en el rendimiento académico.

I. ¿POR QUÉ ESTIMAR LOS COSTOS DE LAS EVALUACIONES?

El principal propósito de este documento es aportar información sobre los costos de las evaluaciones en América Latina. Hasta ahora no hay información publicada sobre el tema. Ha habido quejas anecdóticas sobre los altos costos de las evaluaciones, sobre todo comparados con los limitados fondos discrecionales disponibles en los ministerios y la alta dedicación de tiempo exigida a un puñado de personas calificadas. Por ahora, el debate sobre los costos de evaluación no ha llegado aún a la opinión pública. Esto puede cambiar a medida que las evaluaciones se amplíen y que en una serie de países las agencias externas que actualmente las financian insistan en que estos costos se integren en los presupuestos ordinarios.

EVALUAR ES FUNDAMENTAL

Este trabajo asume que la realización de evaluaciones, al menos a un nivel muestral, es un elemento fundamental de los sistemas educativos modernos, dado que representa una parte

necesaria del proceso de diseño, implementación y evaluación de programas para mejorar la calidad de la educación. El estudio se basa en datos gentilmente puestos a disposición por los funcionarios de cinco países de la región –Chile, Colombia, Honduras, Perú y Uruguay– sobre tres tipos de evaluaciones:

- Universales o censales (es decir, pruebas aplicadas a todos los estudiantes en grados seleccionados) en Chile y Colombia.
- Pruebas muestrales nacionales (es decir, pruebas aplicadas a muestras aleatorias de estudiantes a fin de generalizar resultados para el universo no evaluado) en Colombia, Uruguay, Honduras y Perú.
- Participación en programas internacionales de pruebas muestrales en Colombia, Perú y Uruguay.

Los informantes fueron generalmente los encargados de los programas de evaluación nacional, o colegas por ellos designados. Los costos son comparados con los costos de educar a los estudiantes del nivel en que se efectúa la prueba y como porcentaje de los costos totales por estudiante en cada país. Los datos se presentan en moneda local y en dólares estadounidenses, estimados según su “paridad de poder adquisitivo” (PPA) del año en que se efectuó la prueba³. Cada uno de los países ha seguido su conjunto propio de criterios y definiciones, y no se ha realizado trabajo de campo para confirmar sus estimados. Los datos presentados deben ser, por lo tanto, considerados indicativos más que definitivos. Se espera que estudios futuros puedan calibrar de manera más fina los costos de las evaluaciones en esos y, ojalá, muchos otros países.

3 Este ratio, usado por el Banco Mundial, la Organización para la Cooperación y el Desarrollo Económico (OCDE) y otras agencias internacionales, ha sido diseñado para evitar problemas de sobrevaluación y subvaluación de monedas basadas en tasas de cambio oficiales, tomando en cuenta el poder adquisitivo de las monedas locales, lo que permite que las cifras puedan ser más comparables. Como la moneda en la mayoría de los países en vías de desarrollo está subvaluada, los estimados de la PPA suelen mostrar estimados en US dólares más altos comparados con aquellos estimados basados en tasas de cambio oficiales. Por ejemplo, el ingreso per cápita en Chile y Perú es más del doble de alto si se usa la tasa PPA que la tasa de cambio oficial; el ingreso per cápita en Colombia es tres veces más alto y en Honduras 2,5 veces más. Las tasas PPA y las de cambio oficiales para 2002 pueden encontrarse en el sitio web del Banco Mundial. Ver http://siteresources.worldbank.org/ICPINT/Resources/Table5_7.pdf Todos los valores monetarios de este informe están expresados en equivalencias a dólares americanos convertidos usando PPA.

TRES TIPOS BÁSICOS DE EVALUACIONES

En América Latina, 16 países evalúan hoy a sus estudiantes, ya sea en base a censos o a muestras. Chile evalúa regularmente a todos los estudiantes en grados seleccionados. Otros países de la región que evalúan al universo de estudiantes en grados y años seleccionados son México, Colombia y varios estados del Brasil. El Salvador empezó recientemente a evaluar a todos los estudiantes de grados seleccionados, mientras que Guatemala planea hacerlo en un futuro cercano. Estos países usan pruebas censales para una variedad de propósitos, pero especialmente para diagnóstico y retroalimentación a las partes interesadas. Chile emplea los datos para identificar escuelas y localidades que necesitan ayuda, recompensar a los maestros de buen desempeño, estudiar las causas de los éxitos y fracasos e informar a los padres sobre el progreso de sus hijos. En Chile y México los puntajes relativos de los estudiantes en las evaluaciones tienen efectos sobre los salarios de los maestros o en sus posibilidades de ascenso.

Las pruebas censales pueden usarse para determinar si los alumnos recibirán un diploma o si avanzarán al siguiente nivel educativo. Estas evaluaciones de altas consecuencias se aplican a todos los estudiantes al final de un ciclo, usualmente la educación secundaria. Son comunes en Europa, el Lejano Oriente y en países anglófonos y francófonos, incluido el Caribe de lengua inglesa, como también en algunos estados de Estados Unidos. En América Latina, Costa Rica, República Dominicana y El Salvador exigen hoy una prueba de secundaria que vale entre 20% y 25% de la nota de egreso de secundaria. Los países angloparlantes del Caribe evalúan a todos los estudiantes en un contexto de altas consecuencias (usual, aunque no exclusivamente, para culminar la educación secundaria).

También se pueden evaluar muestras de estudiantes, como es el caso de casi todos los países de la región que tienen programas de evaluación. Las muestras (como también las pruebas censales) pueden usarse para determinar en qué medida el currículo nacional y los estándares acordados están siendo exitosamente implementados en el aula, así como para estudiar cuáles son los factores que contribuyen a los resultados diferenciales de diversas poblaciones de estudiantes. Pueden ayudar a diseñar programas y materiales para

maestros en áreas donde un amplio porcentaje de estudiantes tiene mal rendimiento, y para evaluar el impacto de diversos programas y políticas. Con muestras no es posible identificar, recompensar ni sancionar a escuelas o docentes individuales por su desempeño relativo. En el Perú a veces se han evaluado muestras relativamente grandes para obtener una idea del progreso de unidades territoriales subnacionales, mientras que en Colombia se dispone de resultados que permiten análisis más finos de lo que está sucediendo en algún área metropolitana. En Uruguay el muestreo se usó para evaluar el impacto de la reforma en la capacitación de maestros, la introducción de la escolaridad de jornada completa, y programas dirigidos a niños de las familias más pobres. En Honduras, se ha usado el muestreo para obtener una idea general del aprendizaje en el país, para comparar el aprendizaje por región y ciudad, y para identificar los factores asociados con el aprendizaje. El gobierno federal de Brasil suele evaluar a los estudiantes para comparar el progreso de estados y municipios.

Una variación de las pruebas muestrales se da cuando los países participan en programas internacionales de evaluación, como es cada vez más el caso en América Latina (ver recuadro). Como estas pruebas tienen el propósito de ubicar a un país en el contexto de otros países, se suele requerir una muestra aleatoria relativamente pequeña, de entre 5,000 a 7,000 estudiantes. Este es el mínimo para asegurar un razonable intervalo de confianza en los resultados. En algunos casos se toman muestras de este tamaño en cada una de varias provincias, de manera que cada una es considerada en efecto como un país. En la evaluación PISA 2003 y 2006, México implementó una muestra de más de 30,000 estudiantes a fin de obtener datos confiables para cada estado. Entre los objetivos de participación en pruebas internacionales están el ayudar a establecer nuevas metas nacionales para logros de aprendizaje, reformar el currículo y/o la capacitación de maestros a fin de alcanzar dichas metas. Un resumen completo de la situación de las evaluaciones en América Latina se encuentra en Ferrer (2004).

Participación reciente de América Latina en pruebas internacionales

- Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación- LLECE: 13 países en 1997; 16 países más un estado mexicano en la segunda aplicación en el año 2006.
- El Programa para la Evaluación Internacional de Estudiantes (PISA) de la Organización para la Cooperación Económica y el Desarrollo (OECD's Programme for International Student Assessment): Argentina, Brasil, Chile, México, Perú y Uruguay en el año 2000 y/o 2003; Argentina, Brasil, Chile, Colombia, México y Uruguay en 2006.
- Estudios sobre Tendencias en Matemáticas y Ciencias (TIMMS) (Trends in Mathematics and Science Studies) de la International Association for the Evaluation of Educational Achievement (IEA): Argentina, Chile, Colombia y México en 2003.
- Estudio sobre el Progreso en Lectura y Literacidad (PIRLS) (Progress in Reading and Literacy Study), otra iniciativa apoyada por la IEA: Argentina, Belice y Colombia en el año 2001.

La decisión de si aplicar pruebas censales o muestrales debe tomarse en base al uso cuidadosamente pensado y planificado de los resultados de las pruebas, algo que con frecuencia ha sido pasado por alto en América Latina (ver introducción a Ravela 2001). La tabla 1 presenta un resumen de los objetivos posibles de las evaluaciones, según estos tres tipos de pruebas: participación en programas de evaluación internacional, evaluación de muestras de estudiantes y evaluación de todos los estudiantes en un mismo grado.

SÍNTESIS DEL ESTUDIO

Como veremos, los costos por estudiante difieren significativamente de un país a otro, pero esto no debe interpretarse como que un país está gastando "demasiado" por estudiante en las evaluaciones, sino más bien que las condiciones de evaluación difieren mucho de un país a otro. Por ejemplo, tomar una muestra aleatoria en Uruguay es mucho más simple y menos costoso que tomarla en Perú, cuya gran extensión

TABLA 1

OBJETIVOS POSIBLES DE LAS EVALUACIONES REALIZADAS EN AMÉRICA LATINA

Tipo de prueba	Características	Objetivos posibles
Participación en evaluaciones internacionales	Muestra, generalmente de 5,000 a 7,000 estudiantes.	1. Catalizar el debate nacional sobre el mejoramiento del aprendizaje y sobre la reforma educativa. 2. Evaluar la idoneidad del currículo intencional y del currículo implementado, en términos de su énfasis relativo en aprendizajes de más alta complejidad, comprensión lectora, etc. 3. Evaluar la calidad de la formación inicial y la capacitación en servicio de los maestros y la de otros insumos educativos, incluidos la organización escolar y las prácticas de enseñanza.
Muestra nacional	Muestra de entre 5,000 y 100,000 estudiantes.	Todo lo anterior, así como: 4. Rastrear el progreso a lo largo del tiempo de países, regiones y poblaciones específicas, en dirección a las metas de aprendizaje especificadas. 5. Comparar el desempeño de regiones y diversos grupos poblacionales (urbano/rural, masculino/femenino, identidad étnica, condición socioeconómica) y hallar explicaciones para las diferencias en logros. 6. Seleccionar distritos o grupos poblacionales para esfuerzos de mejoramiento, recompensas y/o sanciones.
Evaluación censal de todos los estudiantes de un grado seleccionado	Universal, usualmente de grados seleccionados y realizadas en años seleccionados, de hasta un millón o más de estudiantes, dependiendo del tamaño del país.	Todo lo anterior, como también: 7. Devolver información a los padres y estudiantes sobre el avance individual en el aprendizaje de un estudiante. 8. Retroalimentar a maestros, escuelas y padres sobre el desempeño de la escuela y del maestro, y ofrecer recompensas y sanciones a fin de estimular los esfuerzos de mejoramiento. 9. Para exámenes de altas consecuencias, certificar que los estudiantes han completado un nivel de educación y que son elegibles para niveles más altos.

física, variedad y dificultad de terrenos elevan significativamente los costos, especialmente de transporte, de un estudio muestral. Como otro ejemplo, una prueba con solo preguntas de opción múltiple, como en el caso de Colombia, es mucho menos costosa de corregir que una con preguntas abiertas. El propósito de este informe no es comparar los altos o bajos costos entre los países, sino más bien brindar un panorama del rango de los costos de evaluación en la región e identificar las situaciones sobre el terre-

no, las alternativas y las decisiones viables (*trade-offs*) que afectan los costos. Determinar de manera más rigurosa los costos que permitan una mayor precisión y la comparabilidad entre países es un desafío a afrontar más adelante y que implicará rastrear los costos por país desde las primeras etapas de las evaluaciones.

Para este estudio, funcionarios de los 5 países recabaron y organizaron los datos sobre costos utilizando una hoja de cálculo estandarizada brin-

dada a ellos por el autor. En algunos casos, los países no brindaron todos los datos solicitados, o los brindaron agregados en varias categorías. Los costos indirectos (*overhead*) estuvieron disponibles en algunos casos y no en otros. Además, algunos costos ocultos, tales como el tiempo que los profesores y los supervisores invierten en administrar las pruebas, no fueron estimados.

Los países suministraron los datos en cuatro rubros generales:

- Preparación de pruebas: análisis del currículo y definición de los objetivos educativos a evaluar, elaboración de los ítemes y del primer borrador de las pruebas, aplicación piloto y retroalimentación, y preparación de la versión final de las pruebas.
- Aplicación de las pruebas: distribución de materiales, preparación del equipo central que supervisará la prueba (incluyendo la contratación de supervisores locales, evaluadores y vigilantes de pruebas) y recolección de datos.
- Procesamiento y análisis: capacitación de personal, procesamiento de preguntas de opción múltiple (por lo general mediante el procesamiento automático de datos), corrección de preguntas abiertas, análisis de resultados y preparación de informes.
- Difusión: impresión y difusión de documentos e informes.

Los estimados suministrados suelen no incluir la capacitación de maestros, el desarrollo del currículo o la revisión de libros de texto en función de los resultados, pero pueden incluir los costos de seminarios breves.

La siguiente sección de este informe resume la información disponible sobre costos aportada por cinco países: Chile, Colombia, Honduras, Perú y Uruguay. Como ya se mencionó, los costos dependen principalmente de las características de las poblaciones muestreadas y de la amplitud y profundidad de la evaluación. Por lo tanto, los lectores deberán revisar el apéndice, que resume las características de los cinco casos nacionales usados en este estudio. La sección III presenta conclusiones tentativas respecto a las evaluaciones de la región, atendiendo a los tres tipos de evaluaciones: pruebas censales nacionales, pruebas muestrales y participación en programas internacionales de evaluación. La última sección discute el impacto potencial so-

bre el aprendizaje que tienen las pruebas, comparado con sus costos.

II. LOS COSTOS DE LAS EVALUACIONES EN AMÉRICA LATINA

PRUEBAS CENSALES

Se obtuvo acceso a datos de dos países que evalúan a todos los estudiantes en determinados grados. Chile aportó datos sobre las pruebas de octavo grado del 2004 en cuatro materias académicas. Colombia aportó datos sobre pruebas efectuadas en el 2003 y el 2004 en el quinto (lenguaje y matemáticas) y noveno grados (ciencias y ciudadanía).

• CHILE

Chile ha estado aplicando las pruebas del SIMCE (Sistema de Medición de la Calidad de la Educación) durante más de 15 años, de manera que ha desarrollado capacidades y un cuadro de expertos tanto dentro como fuera del gobierno. Las pruebas del SIMCE se aplican cada cuatro años al cuarto, octavo y décimo grados⁴. La prueba del 2004 fue rendida por 300,000 estudiantes del octavo grado en 6,500 escuelas y 11,000 aulas en las materias de castellano, matemáticas, ciencias naturales y ciencias sociales. La prueba incluía una mezcla de preguntas con respuestas de opción múltiple y abiertas.

Los datos aportados por Chile son mucho más detallados que los aportados por otros países de la muestra. Pudo desagregar las actividades en preparación, aplicación y análisis de las pruebas. También pudo separar los costos de la corrección de preguntas de opción múltiple (57% del costo de corrección y análisis) de aquellos de la corrección de preguntas abiertas (43%). Chile, como unos cuantos otros países, también aportó alguna información sobre costos indirectos, específicamente, el costo de los salarios de quienes en la unidad de evaluación supervisan todas las actividades de la evaluación. La tabla 2 muestra los costos por categoría tanto en pesos chilenos como en dólares norteamericanos, usando las tasas de cambio PPA.

⁴ Las explicaciones y resultados detallados de las pruebas SIMCE están disponibles en <http://www.simce.cl>

TABLA 2

CHILE - PRUEBAS SIMCE

Algunas especificaciones de la evaluación		
Año	2004	
Áreas de estudio	Lenguaje, matemáticas, ciencias y ciencias sociales	
Tipo de escuela	Pública y privada	
Grados	8º	
Número de estudiantes evaluados	300,000	
Costos de la evaluación	Pesos	US\$ (PPA)
Preparación de la prueba	175,200,000	607,000
Elaboración de los ítemes	76,000,000	263,000
Pilotaje inicial	20,000,000	69,000
Pruebas piloto	30,000,000	104,000
Corrección / Preguntas abiertas	49,200,000	170,000
Aplicación	788,700,000	2,732,000
Diseño y edición	20,000,000	69,000
Impresión de las pruebas	220,000,000	762,000
Impresión de otros materiales	160,000,000	554,000
Distribución	70,000,000	242,000
Aplicación	275,000,000	953,000
Control y supervisión	43,700,000	151,000
Procesamiento y análisis	258,900,000	897,000
Codificación e ingreso de datos	146,400,000	507,000
Preguntas abiertas	112,500,000	390,000
Publicación de resultados	68,250,000	236,000
Informe a cada escuela ^a	68,250,000	236,000
Subtotal	1,291,050,000	4,472,000
Costos indirectos	584,700,000	2,025,000
Personal	540,000,000	1,870,000
Infraestructura	24,000,000	83,000
Equipamiento	7,200,000	25,000
Otros costos generales	13,500,000	47,000
TOTAL	1,875,750,000	6,497,000
Costo por estudiante ^b	4,304	14,90
Costo de educar a un estudiante ^c	519,371	1,799
Costo de evaluación como % del presupuesto total para un grado	0.83	
Costos de evaluación como % del presupuesto total para educación secundaria	0.17	

Fuente: SIMCE

Notas: Tasa de cambio PPA 2002: 288.7 pesos = US\$ 1.

a No se tuvo acceso a los costos de otros documentos, publicaciones y reuniones.

b No incluyendo gastos indirectos.

c Costos por estudiante secundario en el año 2000. Ver Wolff y Gurriá, 2005.

El SIMCE preparó una descripción detallada de cómo calculó los costos para cada rubro, de la siguiente manera:

- Elaboración de ítemes: contrato con una universidad, 200 ítemes validados, incluyendo orientaciones sobre cómo corregir preguntas abiertas.
- Pilotaje inicial: muestra que busca 300 respuestas válidas de estudiantes por ítem.
- Prueba piloto: muestra que busca 3,000 respuestas válidas de estudiantes por ítem, incluyendo el diseño, la impresión y la distribución.
- Diseño y edición: diseño de todos los materiales, incluyendo los cuadernillos de pruebas, listados de las aulas, cuestionarios, formas rotadas, etc.
- Impresión: 300,000 cuadernillos de pruebas para cuatro áreas de aprendizaje, un total de 1.2 millón de cuadernillos.
- Impresión de otros materiales: listado de aulas, encuestas a padres y maestros, un total de 400,000 documentos.
- Distribución: a cerca de 6,000 escuelas, de las cuales 25% eran rurales.
- Aplicación: contratación de 10,000 examinadores.
- Control y supervisión: realizados por personal del SIMCE. Los costos incluyen solo transporte y viáticos.
- Codificación e ingreso digitalizado de datos: codificación automatizada de todos los ítemes y cuestionarios.
- Corrección de preguntas abiertas: corrección de dos secciones completas y 10% de otras 2 secciones, con doble codificación de 10% de las respuestas.
- Informe a las escuelas: resultados personalizados para cada escuela. No incluye otras formas de comunicación.
- Costos indirectos: personal - salarios de 60 miembros del equipo.
- Costos indirectos: infraestructura - espacio físico para el personal del SIMCE.
- Costos indirectos: equipamiento - computadoras.
- Gastos indirectos: otros - gastos generales incluyendo teléfono, electricidad, etc.

El costo total para esta evaluación en particular fue de alrededor de \$ 6.5 millones, o unos \$ 14.90 por estudiante evaluado en octavo grado. Si se añadieran los costos indirectos completos de administrar el programa SIMCE, so-

bre todo los costos salariales del personal del SIMCE dedicado a la administración y supervisión, entonces los costos por estudiante se incrementarían en 45%, a \$ 21.65. Pero como el personal del SIMCE se dedica regularmente a administrar y planificar varias pruebas a la vez, esta cifra podría ser más baja. Sin contar los costos indirectos, el desagregado de costos por actividad es el siguiente: preparación de la prueba, 10,1%; aplicación de la prueba, 63,5%; procesamiento y análisis, 20,9%; difusión, 5,5%. Dentro del rubro preparación de la prueba, preparar y escribir los ítemes (por ejemplo, definir los dominios a ser evaluados y luego escribir los ítemes) representó 60% de los costos. Para el segundo rubro, aplicación de la prueba, la impresión da cuenta de 26% de los costos y la aplicación 36%. Chile incluyó un cuestionario de antecedentes para los estudiantes, maestros y algunos padres, y es probable que esto explique los costos relativamente altos de "otros materiales".

Según UNESCO, el costo estimado de educar a un estudiante de octavo grado en Chile en el año 2000 fue \$ 1,799⁵. Así, el costo de evaluación de \$ 14.90 por estudiante (sin incluir costos indirectos) representaría 0.8% del costo de educar a un estudiante en aquel grado. Al consistir la escuela secundaria de cuatro grados y evaluarse solo un grado en cualquier año dado, se puede decir que el costo de evaluación de la escuela secundaria es equivalente a un cuarto de aquel monto, o 0.17% del gasto público total en educación secundaria.

• COLOMBIA

Desde 1980, el Instituto Colombiano para el Fomento de la Educación Superior (ICFES), agencia pública autónoma, ha estado evaluando a todos los estudiantes que culminan su educación secundaria, como una manera de determinar el acceso a instituciones de educación superior. Estas evaluaciones son similares a las pruebas de admisión del Educational Testing Service (ETS) de Estados Unidos⁶.

5 No se contó con datos de fuentes internacionales para años posteriores.

6 Se puede hallar información sobre las evaluaciones colombianas en <http://www.colombiaaprende.edu.co/html/investigadores/1609/channel.html> y en <http://www.icfes.gov.co>

Hasta hace poco, el gobierno nacional de Colombia aplicaba solo pruebas muestrales en primaria y media⁷. En el 2002 y 2003 empezó a evaluar a todos los estudiantes de quinto y noveno grado cada tres años, vía el programa SABER del Sistema Nacional de Evaluación de la Calidad de la Educación, también encomendado al ICFES. Esta institución ha estado involucrada en pruebas a gran escala desde los años sesenta, tiene competencias profesionales y economías de escala que a menudo faltan en otros países.

La tabla 3 muestra los costos estimados de las pruebas censales de estudiantes en Colombia, tal como lo reportó el personal de ICFES, para la evaluación de todos los estudiantes en quinto y noveno grado en el 2002 y el 2003. En el 2002 se evaluó a los estudiantes en lenguaje y matemáticas y en el 2003 en ciencias y ciudadanía. También se han incluido los costos de administrar las pruebas de admisión a la educación superior para fines comparativos. Los datos aportados por Colombia están incompletos. No se estimaron los costos de procesamiento y análisis de la prueba, ni la difusión para el 2002. Tampoco se estimaron los costos de procesamiento y análisis de la prueba para el 2003. No se estimaron los costos indirectos en ambos años.

En el 2002 se tomaron pruebas de matemáticas y lenguaje a más de un millón de alumnos del quinto y noveno grado a un costo total de \$ 2,5 millones, o un costo por estudiante (de estos grados) de \$ 2,47, sustantivamente más bajo que el costo de \$ 14,90 en Chile. Como se señaló antes, estas cifras no incluyen la corrección ni la difusión de resultados, lo cual posiblemente incrementaría el costo en 25%. No se puede estimar el impacto de los costos indirectos. Dado que Colombia gasta un promedio de \$ 1,365 por estudiante, los costos reportados en el 2002 son equivalentes a 0.2% del costo por estudiante de cada grado evaluado y al 0.02% de los gastos totales en educación, tanto en primaria como en secundaria.

En el 2003 Colombia evaluó a todos los estudiantes en las áreas de ciencia y ciudadanía. Los costos de implementación de estas pruebas

fueron de \$ 5,32 por estudiante evaluado, o más del doble del costo de las pruebas de lenguaje y matemáticas. Llama la atención que menos de 1% del costo corresponde a la preparación y 3% a la difusión.

Como se señaló antes, Colombia ha tenido un programa de larga data para la evaluación de todos los estudiantes del último año de educación secundaria, originalmente aplicado solamente a aquellos que deseaban continuar con la educación superior. Se evalúan hasta nueve materias y cerca de 500,000 estudiantes toman estas pruebas anualmente. El costo estimado es de unos \$ 17,42 por estudiante, significativamente más alto que las pruebas de la educación primaria y media. Las pruebas de salida de secundaria también son totalmente de opción múltiple.

Hay varias razones que explicarían por qué los costos reportados de la evaluación en Colombia son menos de la mitad que los de Chile. A diferencia de lo que sucede en Chile, que depende en mayor grado de consultores, es el mismo personal de ICFES que codifica y analiza las respuestas, lo que hace difícil determinar este costo. ICFES emplea de modo exclusivo ítemes de opción múltiple en lugar de una combinación de esos ítemes con otros de preguntas abiertas. Las pruebas de ICFES no incluyen cuestionarios detallados de antecedentes. Es también posible que las tasas de cambio PPA no capten plenamente las diferencias relativas de salarios entre los países. El gran número de estudiantes evaluados puede permitir a Colombia economías de escala no asequibles a Chile. Además, en el 2002, el ICFES introdujo un sistema de racionalización de costos, que incluyó entrenamiento, computarización de todos los procesos, descentralización e impresión de la mayor parte de materiales en su propia imprenta. Se requeriría mayor revisión e información para confirmar el peso relativo de cada uno de estos factores.

PRUEBAS MUESTRALES

Se pudo obtener datos de pruebas muestrales de Colombia, Honduras, Perú y Uruguay, y los que se describen a continuación. Al revisar esta información, el lector debe tener en cuenta que el Perú es similar en extensión y población a Colombia, mientras que Honduras, aunque de similar tamaño que Uruguay, es un país mucho más pobre que este.

7 La ciudad de Bogotá ha estado evaluando a todos los estudiantes de grados seleccionados (tercero, quinto y noveno) desde 1998.

TABLA 3

COLOMBIA - EVALUACIÓN NACIONAL EN EDUCACIÓN BÁSICA Y EXAMEN DE INGRESO A LA EDUCACIÓN SUPERIOR

Algunas especificaciones de la evaluación	2002		2003		2004	
Áreas de estudio	Lenguaje y matemáticas		Ciencia y ciudadanía		Examen de ingreso a la Educación Superior	
Tipo de escuela	Público y privado		Público y privado		Público y privado	
Grados	5º y 9º		5º y 9º		11º	
Número de estudiantes evaluados	1,030,626		1,034,049		540,716	
Costos de la evaluación	Pesos	US\$ (PPA)	Pesos	US\$ (PPA)	Pesos	US\$ (PPA)
Preparación de la prueba	113,081,000	157,000	29,039,000	38,000	795,127,000	1,002,000
Aplicación de la prueba	1,710,842,000	2,380,000	4,009,667,000	5,298,000	5,878,419,000	7,408,000
Corrección y análisis	8,133,000	11,000	a	a	a	a
Difusión de resultados	ND	ND	121,850,000	161,000	800,126,000	1,008,000
TOTAL	1,832,057,000	2,548,000	4,160,555,000	5,497,000	7,473,672,000	9,418,000
Costo por estudiante	1,778	2,47	4,024	5,32	13,822	17,42
Costo de educar a un estudiante ^b	1,089,948	1,365	1,147,364	1,365	1,203,704	1,516
Costo de evaluación como % del costo total de educar estudiantes en un grado ^c	0.2		0.4			
Costo total de evaluar estudiantes de primaria como % del presupuesto total de educación primaria ^d	0.02		0.04			
Costo total de evaluar estudiantes de secundaria/presupuesto total de educación secundaria ^d	0.02		0.05		0.17	

Fuente: ICFES

Notas: ND = No disponible. No se incluyeron los costos indirectos. Tasa de cambio PPA: en el 2002, 719 pesos = US\$1; en el 2003, 757 pesos = US\$ 1; en el 2004, 794 pesos = US\$ 1.

a Incluido en aplicación de pruebas.

b Promedio ponderado del costo por estudiante de primaria y secundaria en el año 2000, en función al número de estudiantes inscritos en cada nivel. Ver Wolf y Gurriá, 2005.

c Costos totales de evaluación en un grado como porcentaje de gastos totales en ese grado.

d Asume que el número de estudiantes evaluados en cada nivel respeta la distribución de estudiantes entre primaria y secundaria. Los datos son del 2000.

• COLOMBIA

En el 2004 Colombia realizó una prueba muestral nacional del progreso en matemáticas y lenguaje para quinto y noveno grado. La tabla 4 presenta un resumen de esos costos elaborado con la información brindada por el ICFES. No se pudo obtener información sobre los costos del análisis de los datos.

Se evaluó un total de 96,000 estudiantes del quinto y noveno grado, menos de 10% del total de estudiantes en dichos grados. Colombia tomó una muestra relativamente grande de estudiantes, a fin de poder comparar el aprendizaje en centros urbanos y regiones. El costo total

para esta muestra fue equivalente a 40% del costo de las pruebas censales realizadas solo dos años atrás: \$ 1.4 millón, frente a \$ 2.5 millones que costó evaluar a un millón de estudiantes en el 2002. Costó \$ 14,20 evaluar a cada estudiante muestreado, comparado con \$ 2,47 por cada estudiante en la evaluación censal. De otro lado, al evaluarse solo una muestra de estudiantes, el costo del programa de evaluación como porcentaje del presupuesto general fue 0,01%, menos de la mitad del costo de la prueba censal realizada dos años atrás. El costo desagregado de la evaluación muestral fue: preparación de la prueba, 9%; aplicación de la prueba, 78%; análisis de la prueba, no estuvo disponible, y difusión, 13%.

TABLA 4

COLOMBIA - PRUEBA MUESTRAL DE EDUCACIÓN BÁSICA

Algunas especificaciones de la evaluación		
Año	2004	
Áreas de estudio	Lenguaje y matemáticas	
Tipo de escuela	Pública y privada	
Grados	5º y 9º	
Número de estudiantes evaluados	96,242	
Costos de la evaluación	Pesos	US\$ (PPA)
Preparación de la prueba	100,000,000	126,000
Aplicación	838,318,000	1,056,000
Corrección y análisis	ND	ND
Difusión de resultados	159,414,000	181,000
TOTAL	1,097,732,000	1,363,000
Costo por estudiante	11,405	14,20
Costo de educar a un estudiante ^a	1,203,019	1,365
Costo total de evaluar a estudiantes de primaria como porcentaje del presupuesto total de educación primaria ^{b,c}	0.01	
Costo total de evaluación de estudiantes de secundaria como porcentaje del presupuesto total de educación secundaria ^c	0.01	

Fuente: ICFES

Notas: ND = No disponible. No se incluyeron los costos indirectos. Tasa de cambio PPA: en el 2002: 794 pesos colombianos = US\$ 1

a Promedio ponderado del costo por estudiante de primaria y secundaria en el año 2000, en función al número de estudiantes inscritos en cada nivel. Ver Wolff y Gurriá, 2005.

b Calculado en base a los costos totales de la evaluación de un grado como porcentaje de los gastos totales en educación primaria, y no en el costo de evaluación por estudiante, ya que solo se evaluó una muestra de estudiantes.

c Se asume que el número de estudiantes evaluados en cada nivel respeta la distribución de estudiantes entre primaria y secundaria; los datos son de 2000.

• URUGUAY

Se dispuso de datos sobre las pruebas muestrales de la Administración Nacional de Educación Pública (ANEP) de Uruguay para los años 2001, 2002 y 2003, en preescolar y en primero, segundo, sexto y duodécimo grado. En el 2002, Uruguay evaluó en las áreas de lenguaje y matemáticas a todas las escuelas que funcionaban con el nuevo programa de jornada completa, y también tomó una muestra aleatoria de otras escuelas. Se evaluó a un total de 9,171 estudiantes. En el 2003 se evaluó a 12,993 estudiantes del nivel secundario, superior o diversificado –una muestra en las escuelas convencionales y todos los estudiantes de las escuelas técnicas– en las áreas de lenguaje, matemáticas, ciencias sociales y ciencias naturales. También se administró cuestionarios para recoger otros datos contextuales a estudiantes, maestros, directores de escuela y padres de familia⁸. Los costos reportados fueron los que aparecen en la tabla 5:

El costo total de las evaluaciones fue \$ 111,000 en el 2002 y \$ 266,000 en el 2003. El costo por estudiante en el 2002 para los alumnos de sexto grado que tomaron dos pruebas fue de \$ 12. El costo por estudiante en el 2003 fue más alto (\$ 20), tal vez como resultado de que se evaluaron cuatro materias en vez de dos, y también de que evaluar a alumnos del duodécimo grado requiere medir capacidades de mayor orden de complejidad que evaluar a los de sexto grado. El costo por estudiante para las pruebas de 2001 fue aún mayor, debido a lo pequeño de la muestra y a que se evaluaron tres grados a la vez. Al tener Uruguay un sistema y presupuestos educativos mucho más pequeños que los de Colombia y evaluar un mayor porcentaje de sus estudiantes, el costo en términos de porcentaje del presupuesto nacional fue mucho mayor que el de Colombia: 0.03% en el año 2001 y 0.07% en el 2002, comparados a 0.01%. En el 2002, el desagregado de costos por tipo de actividad fue el siguiente: elaboración de la prueba, 10.1%; aplicación, 46.9%; análisis, 11.7%; y difusión, 20.7%. Uruguay prestó gran atención a la difusión de los resultados de la evaluación de 2002.

• PERÚ

En 2001 Perú evaluó a 34,000 estudiantes de cuarto y sexto de primaria y cuarto y quinto grado de secundaria en matemáticas y comunicación. En 2004 emprendió un programa muy ambicioso de pruebas, evaluando a un total de 70,000 estudiantes de segundo y sexto grado de primaria, como también de tercero y quinto de secundaria, o 17,500 alumnos en cada grado. Las pruebas cubrieron matemáticas y comunicación en todos esos grados, ciudadanía en sexto de primaria y quinto de secundaria. En el caso de lenguaje en sexto grado, también se aplicaron pruebas en dos lenguas indígenas. La tabla 6 resume los costos del programa.

El costo total de evaluación en 2001 fue \$1.7 millón; en 2004, \$ 4.9 millones, más del doble del monto que en el 2001. En 2004 se incluyó una tercera materia (ciudadanía), se evaluó a más del doble de número de estudiantes, y se reportó el control de campo para asegurar condiciones confiables y adecuadas para la aplicación de pruebas. El costo promedio de evaluación por estudiante en el 2001 fue \$ 52 y \$ 70 en el 2004.

Las evaluaciones en Perú, sobre todo en el 2004, fueron integrales y complejas. Se evaluaron cuatro grados y tres materias, frente a uno o dos grados en el caso de los otros países que proporcionaron datos. También se tomaron pruebas en lenguas indígenas extrayendo una muestra especial de escuelas bilingües. Se administraron cuestionarios a estudiantes, padres y maestros. Los maestros recibieron cuestionarios sobre “oportunidades de aprendizaje”, que preguntaban sobre tópico de matemáticas y lenguaje que ellos realmente habían enseñado, y también se evaluaron sus propios conocimientos de estas áreas. Se contrataron a terceros para la aplicación y corrección de las pruebas. Las preguntas abiertas a los estudiantes representaron entre 30% y 100% de los ítems y fueron calificados dos veces para asegurar la confiabilidad. La prueba aplicada a los maestros fue totalmente abierta. La aplicación de la prueba tomó 5 días. Se brindó capacitación de 4 a 5 días a los aplicadores de los instrumentos. La implementación de esta prueba era la principal actividad de la Unidad de Medición de la Calidad Educativa. En ese tiempo, Perú estaba empezando su sistema de evaluación, no

⁸ Se puede encontrar información sobre el programa de evaluación de Uruguay en <http://www.anep.edu.uy>

contaba con un marco institucional y dependía mucho de los consultores. Tanto el Banco Mundial como el Banco Interamericano de Desarrollo aportaron financiamiento parcial. El costo

total de la evaluación en el año 2004 que fue de \$ 4.9 millones, fue equivalente a solo 0.15% de los gastos en educación primaria y 0.13% de los gastos en educación secundaria.

TABLA 5

URUGUAY - PRUEBA MUESTRAL DE EDUCACIÓN PRIMARIA Y MEDIA

Algunas especificaciones de la evaluación	2001		2002		2003	
Áreas de estudio	Lenguaje, matemáticas, desarrollo cognitivo y afectivo		Lenguaje y matemáticas		Lenguaje, matemáticas, ciencias sociales y ciencias naturales	
Tipo de escuela	Pública y privada		Pública y privada		Pública y privada	
Grados	Preescolar, 1º y 2º		6º		12º	
Número de estudiantes evaluados	2,387		9,171		12,993	
Costos de la evaluación	Pesos	US \$ (PPA)	Pesos	US \$ (PPA)	Pesos	US \$ (PPA)
Preparación de la prueba	299,000	35,000	229,000	23,000	608,000	54,000
Aplicación de la prueba	464,000	55,000	524,000	52,000	1,618,000	144,000
Corrección y análisis de la prueba	36,000	4,000	134,000	13,000	742,000	66,000
Difusión de resultados	64,000	8,000	226,000	23,000	15,000	2,000
TOTAL	863,000	102,000	1,113,000	111,000	2,983,000	266,000
Costo por estudiante	362	43	121	12	230	20
Costo de educación de un estudiante	8,592	1,016	10,104	1,011	13,657	1,219
Costo total de evaluar a estudiantes de primaria como porcentaje del presupuesto total de educación primaria ^a	0.03		0.03			
Costo total de evaluación de estudiantes de secundaria como porcentaje del presupuesto total de educación secundaria					0.07	

Fuente: ANEP.

Notas: Las tasas de cambio PPA variaron cada año. En 2003 la tasa fue 11.21 pesos=US\$ 1. No incluye el costo del reducido número de personal que trabaja a tiempo completo en todos los estudios. Este grupo, como también otras instituciones, conduce ciertas actividades de difusión que tampoco han sido costeadas.

^a Calculado en base a los costos totales de la evaluación de un grado como porcentaje de los gastos totales en educación primaria, y no en el costo de evaluación por estudiante, ya que solo se evaluó una muestra de estudiantes.

TABLA 6

PERÚ - PRUEBA MUESTRAL DE EDUCACIÓN PRIMARIA Y SECUNDARIA

Algunas especificaciones de la evaluación	2001		2004	
Áreas de estudio	Matemáticas y comunicación		Matemáticas, comunicación y ciudadanía	
	Muestra en escuelas bilingües: Comprensión de lectura y escritura en primaria y segundas lenguas (aymara y quechua / castellano)		Muestra en escuelas bilingües: Comprensión de lectura y escritura en primaria y segundas lenguas (aymara y quechua / castellano)	
Tipo de escuela	Pública y privada		Pública y privada	
Grados	4º, 6º, 10º		2º, 6º, 9º y 11º	
Número de estudiantes evaluados	34,000		70,000	
Costos de la evaluación	Soles	US\$(PPA)	Soles	US\$ (PPA)
Preparación de la prueba	250,000	166,000	1,142,000	760,000
Aplicación de la prueba	1,474,000	980,000	5,083,000	3,384,000
Corrección y análisis de la prueba	587,000	390,000	899,000	598,000
Difusión de resultados	331,000	220,000	214,000	142,000
TOTAL	2,642,000	1,757,000	7,338,000	4,885,000
Costo por estudiante	78	52	105	70
Costo de educar a un estudiante ^a	751	500	751	500
Costo total de evaluar a estudiantes de primaria como porcentaje del presupuesto total de educación primaria ^b	0.06		0.15	
Costo total de evaluación de estudiantes de secundaria como porcentaje del presupuesto total de educación secundaria	0.05		0.13	

Fuente: UMC

Notas: No incluye costos de *overhead*. Tasa de cambio PPA 2001/2004: 1.5 soles peruanos=US\$ 1.

a Promedio ponderado del costo por estudiante de primaria y secundaria en el año 2000, en función al número de estudiantes inscritos en cada nivel. Ver Wolff y Gurriá, 2005.

b Calculado en base a los costos totales de la evaluación de un grado como porcentaje de los gastos totales en educación primaria, y no en el costo de evaluación por estudiante, ya que solo se evaluó una muestra de estudiantes.

• HONDURAS

En el 2002 y el 2004, Honduras evaluó el rendimiento de más de 40,000 estudiantes de tercero y sexto grado en lenguaje, matemáticas y ciencias⁹. La tabla 7 presenta un resumen de los costos aportados por consultores de la Unidad de Medición de la Calidad Educativa (UMCE), localizada en la Universidad Pedagógica Nacional Francisco Morazán de Honduras.

Honduras evaluó a 24-25,000 estudiantes de tercer grado y entre 18,000 y 20,000 estudiantes de sexto grado, de un total de 140,000 alumnos estimados en cada grado. Los costos totales fueron \$ 1.9 millón en el 2002 y \$ 2.3 millones en el 2004. El costo por estudiante evaluado fue \$ 46 en el 2002 y algo más de \$ 50 en el 2004. Los costos totales de evaluación representaron entre 0.28% y 0.33% del presu-

puesto nacional asignado a la educación primaria. Este es un monto significativo, si se considera que más de 90% del presupuesto de Honduras se destina a salarios de maestros y que el número de estudiantes evaluados es relativamente pequeño. El desagregado de los gastos por actividad en el 2004 fue reportado de la siguiente manera: preparación de la prueba, 20%; aplicación, 50%; procesamiento y análisis, 15%; y difusión, 15%. Honduras tiene un número muy pequeño de expertos en evaluación. Esta escasez condujo probablemente a apoyarse en consultores externos, lo que elevó los costos. Otros elementos del programa hondureño que impactaron sobre los costos son que se evaluaron múltiples grados (2) y materias (3), se realizó un muestreo y análisis de datos más cuidadoso que en años anteriores; el financiamiento externo conllevó regulaciones procedimentales que pueden haber generado

TABLA 7

HONDURAS - EVALUACIÓN MUESTRAL

Algunas especificaciones de la evaluación	2002		2004	
Áreas de estudio	Lenguaje, matemáticas y ciencias naturales		Lenguaje, matemáticas y ciencias naturales	
Tipo de escuela	Pública y privada		Pública y privada	
Grados	3º y 6º		3º y 6º	
Número de estudiantes evaluados	42,572		45,657	
Costos de la evaluación	Lempiras	US\$ (PPA)	Lempiras	US\$ (PPA)
Preparación de la prueba	2,377,000	391,000	3,081,000	458,000
Aplicación de la prueba	5,942,000	977,000	7,703,000	1,146,000
Corrección y análisis de la prueba	1,783,000	293,000	2,311,000	344,000
Difusión de resultados	1,783,000	293,000	2,311,000	344,000
TOTAL	11,885,000	1,953,000	15,406,000	2,292,000
Costo por estudiante	279	45,88	337	50,20
Costo de educación de un estudiante ^a	2,080	342	2,299	342
Costo total de la evaluación/gastos				
totales en educación primaria ^b	0.28		0.33	

Fuente: UMC.

Notas: No incluye costos de *overhead*. Tasa de cambio PPA 2002: 6.1 lempiras = US\$ 1; 2004: 6.6 lempiras = US\$ 1.

a Promedio ponderado del costo por estudiante de primaria y secundaria en el año 2000, en función al número de estudiantes inscritos en cada nivel. Ver Wolff y Gurriá, 2005.

b Calculado en base a los costos totales de la evaluación de un grado como porcentaje de los gastos totales en educación primaria, y no en el costo de evaluación por estudiante, ya que solo se evaluó una muestra de estudiantes.

9 Puede encontrarse información sobre la evaluación de Honduras en <http://www.upnfm.edu.hn/proyectos/umce/acercade.htm>

costos más altos y el que la unidad de evaluación estuviera localizada en la Universidad Pedagógica en vez del Ministerio de Educación, lo que se hizo en parte para asegurar una mejor calidad técnica.

PARTICIPACIÓN EN EVALUACIONES INTERNACIONALES

Los países latinoamericanos participan cada vez más en programas internacionales de evaluación. Seis países (Argentina, Brasil, Chile, México, Perú y Uruguay) han participado en el esfuerzo internacional de evaluación PISA, que mide capacidades en lectura y razonamiento matemático y científico de estudiantes de 15 años. La prueba incluye una combinación de preguntas abiertas y de opción múltiple, como también cuestionarios sobre antecedentes. La prueba PISA exige un mínimo de 5,000 estudiantes escogidos al azar de todo un país para asegurar resultados estadísticamente válidos. En el 2003 requería a cada país pagar más de \$ 100,000 para participar en el programa.

Se cuenta con datos de los costos de PISA de Uruguay para el 2003 y de Perú para el 2000. También se cuenta con datos sobre la participación de Colombia en la Prueba Internacional de Comprensión Lectora (PIRLS), que mide las habilidades de lectura de estudiantes de cuarto grado, y que normalmente requiere a los países participantes un pago anual de \$ 20,000 durante tres años.

• URUGUAY

Los costos de la participación de Uruguay en la prueba PISA del 2003 se resumen en la tabla 8:

El costo total de la participación de Uruguay fue \$ 311,000. La mayor parte de este gasto fue el pago de \$ 110,000 a la oficina central de PISA para su administración central, lo que incluye servicios de la preparación de las pruebas, orientación para las aplicaciones, viajes internacionales y análisis. Uruguay evaluó un total de 5,797 estudiantes. Desde que Uruguay realizó la evaluación ha publicado varios análisis basa-

TABLA 8
URUGUAY - PARTICIPACIÓN EN PISA

Algunas especificaciones de la evaluación		
Año	2003	
Áreas de estudio	Lectura, matemáticas, ciencias, solución de problemas	
Tipo de escuela	Pública y privada	
Grados	Estudiantes de 15 años (principalmente 10º grado)	
Número de estudiantes evaluados	5,797	
Costos de la evaluación	Pesos	US\$ (PPP)
Costo de participación en PISA/OECD	1,220,000	110,000
Preparación de la prueba	349,000	35,000
Aplicación de la prueba	839,000	75,000
Corrección y análisis de la prueba	645,000	58,000
Difusión de resultados	398,000	34,000
TOTAL	3,451,000	311,000
Costo por estudiante	595	53,64
Costo de educar a un estudiante ^a	13,530	1,219
Costo total de evaluación/presupuesto total de educación secundaria		0.08

Fuente: ANEP.

Notas: La tasa en el 2003: 11.21 pesos = US\$ 1. Las tasas de cambio PPA variaron cada año. No incluye el costo del reducido número de personal que trabaja a tiempo completo en todos los estudios. Este grupo, como también otras instituciones, conduce ciertas actividades de difusión que tampoco han sido costeadas.

a El costo de educar a un estudiante corresponde al costo por estudiante a nivel secundario en el año 2000. Ver Wolf y Gurriá (2005).

dos en PISA. Esto puede no haber sido incluido en los costos de difusión. Los costos desagregados fueron: participación en PISA, 35%; preparación de la prueba, 11%; aplicación de la prueba, 24%; análisis, 19%; y difusión 11%.

• PERÚ

La tabla 9 resume los costos de la participación de Perú en PISA 2000.

La preparación de la evaluación costó \$ 171,000. Esto incluyó el diseño de la muestra y los ítemes, los viajes a las reuniones internacionales de PISA, y la revisión de las traducciones al español. El costo más alto de Perú frente a los de Uruguay es resultado de las dificultades de realizar pruebas a muestras aleatorias en un país extenso y montañoso.

Perú gastó solamente \$ 33,000 en difusión. En cualquier caso, el costo total de \$ 480,000 en la prueba PISA fue solo 0.04% de los gastos anuales en la educación secundaria del Perú. Los costos desagregados por actividad fueron los siguientes: membresía en PISA, 21%; preparación, 36%; aplicación, 27%; procesamiento y análisis, 10%; y difusión, 7%.

• COLOMBIA

El costo de participación de Colombia en el estudio PIRLS de competencia en lectura en cuarto grado en el 2001 es similar al de Perú para la prueba PISA, como puede verse en la tabla 10, a pesar de que Colombia no reportó costos de análisis y difusión. El costo de aplicación de la prueba fue significativamente más alto que el del Perú.

TABLA 9

PERÚ - PARTICIPACIÓN EN PISA

Algunas especificaciones de la evaluación		
Año	2001	
Áreas de estudio	Lectura, matemáticas y ciencias	
Tipo de escuela	Pública y privada	
Grados	Estudiantes de 15 años (principalmente de 9º y 10º grado)	
Número de estudiantes evaluados	5,190	
Costos de la evaluación	Soles	US\$ (PPA)
Costo de participación	150,000	100,000
Preparación de la prueba	257,000	171,000
Aplicación de la prueba	193,000	128,000
Corrección y análisis de la prueba	73,000	49,000
Difusión de resultados	49,000	33,000
TOTAL	722,000	480,000
Costo por estudiante	139	92
Costo de educar a un estudiante ^a	832	553
Costo total de la evaluación/gastos totales en educación secundaria		0.04

Fuente: UMC

Notas: La tasa en el 2000 fue de 1.5 soles=US\$ 1. No se incluyen los costos del personal de la UMC que participó en la preparación de la evaluación y la corrección y análisis de la prueba.

^a El costo de educar a un estudiante corresponde al costo por estudiante a nivel secundario en el año 2000. Ver Wolf y Gurría (2005).

TABLA 10

COLOMBIA - COSTOS DE PARTICIPACIÓN EN PIRLS

Algunas especificaciones de la evaluación		
Año	2001	
Áreas de estudio	Lectura	
Tipo de escuela	Pública y privada	
Grados	4º	
Número de estudiantes evaluados	5,131	
Costos de la evaluación	Pesos	US\$ PPA
Costo de asociación a la International Evaluation Association	41,220,000	60,000
Preparación de la prueba	93,986,000	137,000
Aplicación de la prueba	276,582,000	402,000
Corrección y análisis de la prueba	N.D.	N.D.
Difusión de resultados	N.D.	N.D.
TOTAL	411,788,000	599,000
Costo por estudiante	80,255	117
Costo de educar a un estudiante ^a	840,228	1,223
Costo total de evaluación/gastos totales en educación primaria	0.01	

Fuente: ICFES.

Notas: ND= No Disponible. No incluye costos indirectos. La tasa de cambio PPA en el 2001 fue 688 pesos colombianos = US\$ 1

^a El costo de educar a un estudiante corresponde al costo por estudiante a nivel primario en el año 2000. Ver Wolff y Gurriá, 2005.

Costo de participar en la prueba regional de UNESCO

También se encuentra disponible el costo de participación en el programa de evaluación regional de la UNESCO (Laboratorio Latinoamericano de Evaluación de la Calidad en Educación, LLECE), realizado en 1997. Este programa evaluó a alumnos de tercero y cuarto grado en 1997 y 1998 en lenguaje y matemáticas en 13 países. La tabla 11 presenta un estimado de los costos de administración central, como también los de la participación de los 13 países:

El costo total de todo el programa fue algo menos de \$2 millones de dólares, un costo aproximado de \$166,000 por país, bastante menos que la participación en PISA o PIRLS. Los costos más bajos son resultado, en parte, de que cada país solo tuvo que aportar \$10,000 anuales a UNESCO, comparado con los 30 a 50,000 dólares por año por participar en PISA, TIMSS o PIRLS. Los viajes para las reuniones de expertos para un estudio regional son mucho menos cos-

TABLA 11

COSTOS ESTIMADOS DE LA PRUEBA LLECE DE UNESCO 1997-99

Fuentes	US\$
Apoyo directo de los países	269,000
Apoyo de UNESCO	80,000
Fundación Ford	50,000
BID	500,000
Subtotal	899,000
Costos internos en los países ^a	1,100,000
TOTAL	1,999,000

Fuente: estimados del BID en base a documentación aportada por UNESCO.

^a Los costos son estimados para los 12 países

tosos que para uno internacional. Casi todos los ítemes fueron de opción múltiple. El tamaño de la muestra en cada país fue de alrededor de 5,000 estudiantes. Esta ronda de pruebas fue criticada por tener una serie de inconsistencias técnicas, que pueden haber sido resultado en parte de un inadecuado financiamiento para especialistas. Se anticipa que los costos para la nueva ronda de evaluaciones de LLECE serán significativamente más altos que la primera ronda. LLECE está manteniendo sus costos bajos gracias al empleo de expertos latinoamericanos de las instituciones evaluadoras de los países que están realizando la mayor parte del trabajo preparatorio, pero también ha establecido un comité internacional de supervisión para asegurar la calidad técnica. Una estimación preliminar de los costos totales, reportados por el equipo de LLECE, rodea los \$ 5.2 millones para 17 países y un Estado, o cerca de \$ 307 mil por participante.

III. CONCLUSIONES SOBRE LOS COSTOS DE LAS EVALUACIONES EN AMÉRICA LATINA

La tabla 12 y los gráficos siguientes resumen la información recogida sobre el costo de las evaluaciones. Esta sección resalta algunas conclusiones que se pueden derivar sobre los costos de la evaluación en América Latina (Ver tabla 12).

Basándonos en esta tabla y en los gráficos que posteriormente serán presentados, ¿qué se puede concluir acerca de los costos de las evaluaciones en América Latina?

EL COSTO DE LAS EVALUACIONES COMO PORCENTAJE DE LOS COSTOS DE LA EDUCACIÓN

Como se observa en el gráfico 1, las evaluaciones, tal como se vienen realizando en la región, no son una carga financiera significativa, representando bastante menos del 1% del presupuesto total del nivel de estudio (primaria o secundaria).

Los costos de participación en los programas internacionales, que suelen exigir muestras de cerca de 5,000, fluctúan entre \$ 300,000 y \$ 600,000 en los países estudiados. En Colombia, las evaluaciones alcanzan 0.02% o menos del presupuesto nacional para la educación primaria o secundaria. En la mayoría de los demás países los costos de evaluaciones varían

entre 0.02% y 0.17% del presupuesto nacional. Honduras, probablemente por su necesidad de depender de consultores externos y por su decisión de tercerizar todos los aspectos del proceso de evaluación, muestra los costos relativos más altos, que todavía son solo el 0.3% del presupuesto público para la escuela primaria. Resulta evidente que implementar una prueba para todos los estudiantes de un grado requiere desembolsos mayores que evaluar solo una muestra, como se argumenta más adelante. Por ejemplo, Colombia gastó \$ 2.5 millones en evaluar a todos los estudiantes de un grado, pero \$ 1.2 millón para evaluar una muestra amplia en el 2004. Es importante señalar que ningún país de la región evalúa a todos los estudiantes de todos los grados.

Si bien ninguno de estos montos es extraordinariamente grande, debe recordarse que los presupuestos educativos y del sector público en América Latina son bastante pequeños, debido tanto al reducido PBI como a la baja recaudación tributaria, así como también a una falta de claridad en las prioridades. Los salarios de los maestros y del personal administrativo tienden a absorber la parte mayor de esos recursos: entre 70% y 91% de los gastos totales (Bruneforth *et al.* 2004), y los presupuestos “discrecionales” son bastante limitados, una situación que no es fácil de resolver en el corto plazo. Es probable que la evaluación sea percibida como una carga mayor en algunos países pequeños y pobres (por ejemplo, aquellos con un PBI per cápita menor a \$ 5,000) con limitada capacidad de recursos presupuestarios y humanos como es el caso de Honduras, a diferencia de los países más grandes y avanzados económicamente.

De otro lado, las evaluaciones son baratas si se las compara con la mayoría de intervenciones propuestas para mejorar el aprendizaje. Un estimado reciente de costos de la educación primaria da una idea de los costos de evaluación comparados con otras intervenciones que suelen ser discutidas e implementadas en América Latina (Ver la tabla 13).

La evaluación es una de las innovaciones menos costosas en la reforma de la educación primaria requiriendo apenas un 1% del costo de incrementar los salarios de los maestros, reducir el tamaño de las clases y fortalecer la capacitación de los maestros. Esto sugiere que evaluar puede ser

TABLA 12

RESUMEN DE LOS COSTOS DE LAS EVALUACIONES DE AMÉRICA LATINA

País	Tipo de evaluación	Año	Áreas de estudio	Grados	Costo total de la evaluación (US\$ PPA)	Número de estudiantes evaluados	Costo por estudiante (US\$ PPA)	Costos de la evaluación como % del presupuesto general correspondiente al nivel educativo ^a	
								Primaria	Secundaria
Chile	Censal	2004	Lenguaje, matemáticas, ciencias y ciencias sociales	8°	4,472,000	300,000	15		0.17
Colombia	Internacional	2001	Lectura	4°	599,000	5,131	117	0.01	
	Censal	2002	Lenguaje y matemáticas	5° y 9°	2,548,000	1,030,626	2	0.02	0.02
	Censal	2003	Ciencia y Ciudadanía	5° y 9°	5,497,000	1,034,049	5	0.04	0.05
	Censal	2004	Examen de ingreso a Educación Superior	11°	9,418,000	540,716	17		0.17
	Muestral	2004	Lenguaje y matemáticas	5° y 9°	1,363,000	96,242	14	0.01	0.01
Honduras	Muestral	2002	Lenguaje, matemáticas y ciencias naturales	3° y 6°	1,953,000	42,572	46	0.28	
	Muestral	2004	Lenguaje, matemáticas y ciencias naturales	3° y 6°	2,292,000	45,657	50	0.33	
Perú	Internacional	2001	Lectura, matemáticas y ciencia	Alumnos de 15 años en 7° o posterior	480	5,190,000	92		0.04
	Muestral	2001	Matemáticas y comunicación	4°, 6°, 10°	1,757	34,000,000	52	0.06	0.05
	Muestral	2004	Matemáticas, y lectura y escritura	2°, 6°, 9° y 11°	4,885	70,000,000	70	0.15	0.13
	Internacional	2003	Lectura, matemáticas, ciencia y solución de problemas	Alumnos de 15 años en 7° o posterior	311	5,797,000	54		0.08
Uruguay	Muestral	2001	Lenguaje, matemáticas, desarrollo cognitivo y afectivo	Preescolar, 1° y 2°	102	2,387,000	43	0.03	
	Muestral	2002	Lenguaje y matemáticas	6°	111	9,171,000	12	0.03	
	Sample	2003	Lenguaje, matemáticas, ciencias sociales y ciencias naturales	12°	266	12,993,000	20		0.07

^a El presupuesto total para educación por nivel corresponde a sus respectivos valores en el 2000. Cuando se aplican pruebas tanto a primaria como a secundaria se asume la mitad del costo por evaluación a cada nivel. Los costos indirectos no están incluidos. Para posibles explicaciones sobre diferencias entre el costo total absoluto y el costo unitario de esta tabla, ver la sección Variabilidad de los costos de la evaluación líneas abajo.

GRÁFICO 1

COSTOS DE EVALUACIÓN COMO PORCENTAJE DEL PRESUPUESTO TOTAL DEL NIVEL EDUCATIVO CORRESPONDIENTE

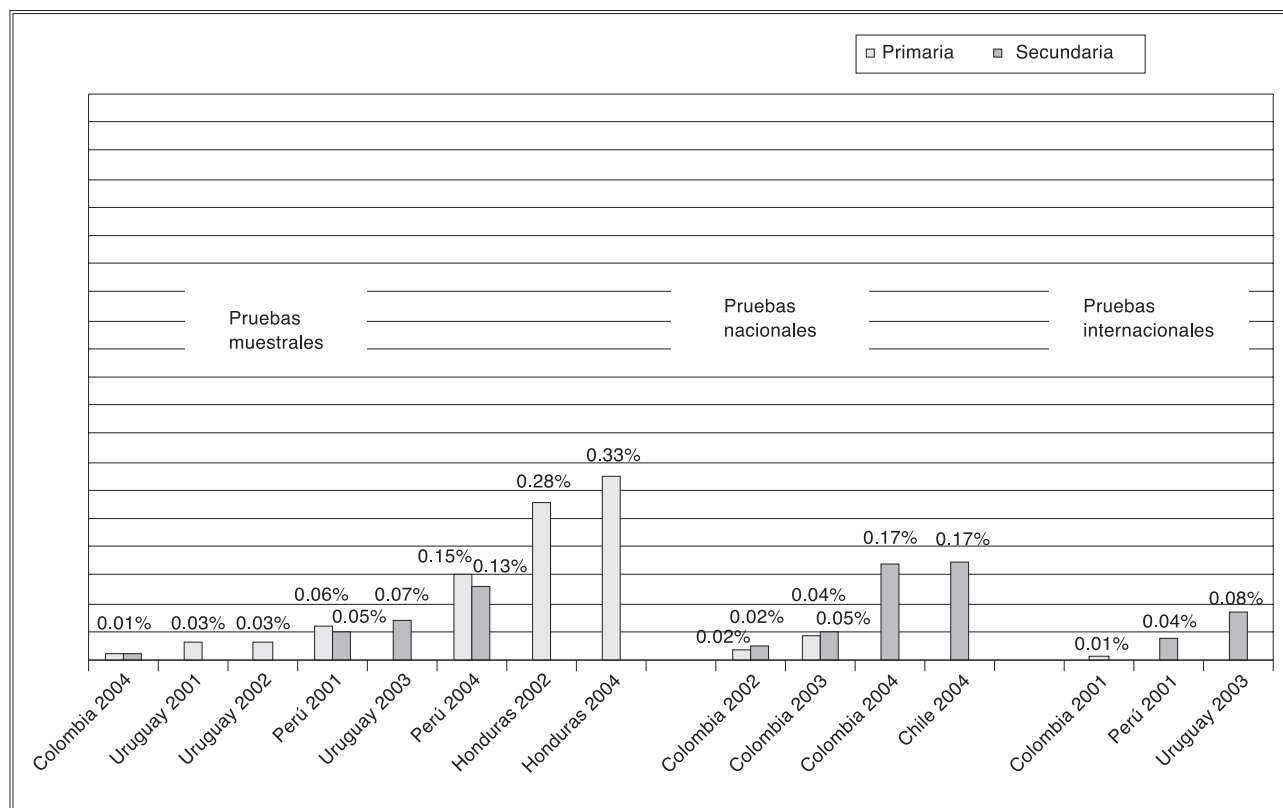


TABLA 13

COSTO ESTIMADO DE INTERVENCIONES POTENCIALES EN EDUCACIÓN PRIMARIA

Tipo de intervención	Incremento estimado de diversos costos unitarios (%)
Evaluar una muestra de estudiantes de cuarto grado	0.01 a 0.3
Evaluar a todos los alumnos de cuarto grado	0.04 a 0.2
Reducir el tamaño de la clase en 10%	9
Preescolaridad para todos los niños en riesgo (50%)	8.3
Elevar salarios de maestros en 20%	18
Capacitación para maestros focalizada (una semana por año)	2.3
Instrucción interactiva por radio	0.5

Fuente: Adaptado de Schiefelbein, Wolff y Schiefelbein (2000)

un complemento relativamente económico para los esfuerzos de reforma, siempre y cuando los resultados sean técnicamente adecuados y utilizados para la toma de decisiones. Por supuesto, en el caso de que las evaluaciones tengan una calidad técnica tan baja que no permitan llegar a ninguna conclusión válida acerca del aprendizaje, o si no se realiza una adecuada diseminación de los resultados o estos no son utilizados, resultarán ciertamente muy costosas. A la vez, existen costos ocultos en las evaluaciones que no han podido ser tomados en cuenta aquí. Además de conseguir mejores estimados de los costos indirectos, una contabilización total de costos debería tomar en consideración el tiempo que tardan los maestros en algunos países para aplicar pruebas, recoger los materiales y entregar los resultados a los supervisores, lo cual podría fácilmente equivaler a un día completo de escolaridad. Si todos los estudiantes de un determinado grado fueran evaluados, ello representaría cerca de 0.55% de los costos totales de aquel grado, suponiendo un año escolar de 180 días, mucho más que los costos de evaluación reportado. También significaría menos tiempo para la enseñanza. Además es posible que los supervisores escolares pasen muchos días preparando e implementando las pruebas, sin ningún reembolso directo y, como ya se señaló, hay costos adicionales asociados con la reforma curricular, la capacitación y otros programas que se derivan, o deberían derivarse, de las evaluaciones.

VARIABILIDAD DE LOS COSTOS DE LAS EVALUACIONES

Los costos de evaluación varían mucho de un país a otro, aun cuando ellos emprendan programas similares. Por ejemplo, Colombia evaluó a 96,000 estudiantes a un costo de \$ 1.4 millón, mientras que Honduras evaluó a 46,000 estudiantes con \$ 2.3 millones, y Perú a 70,000 estudiantes con \$ 4.9 millones. La participación en PISA costó \$ 480,000 en Perú y \$ 311,000 en Uruguay, mientras que la participación de Colombia en PIRLS costó \$599,000. Las diferencias de costos entre los países pueden ser explicadas en parte por diferencias en el número de grados y materias evaluados. El gráfico 2 muestra (1) los costos totales de evaluar para cada país en un año particular y para un tipo particular de evaluación (por ejemplo, censal, muestral o participación en una prueba internacional); y (2) cuál sería el costo de evaluar solo un grado y una materia (costo ajustado).

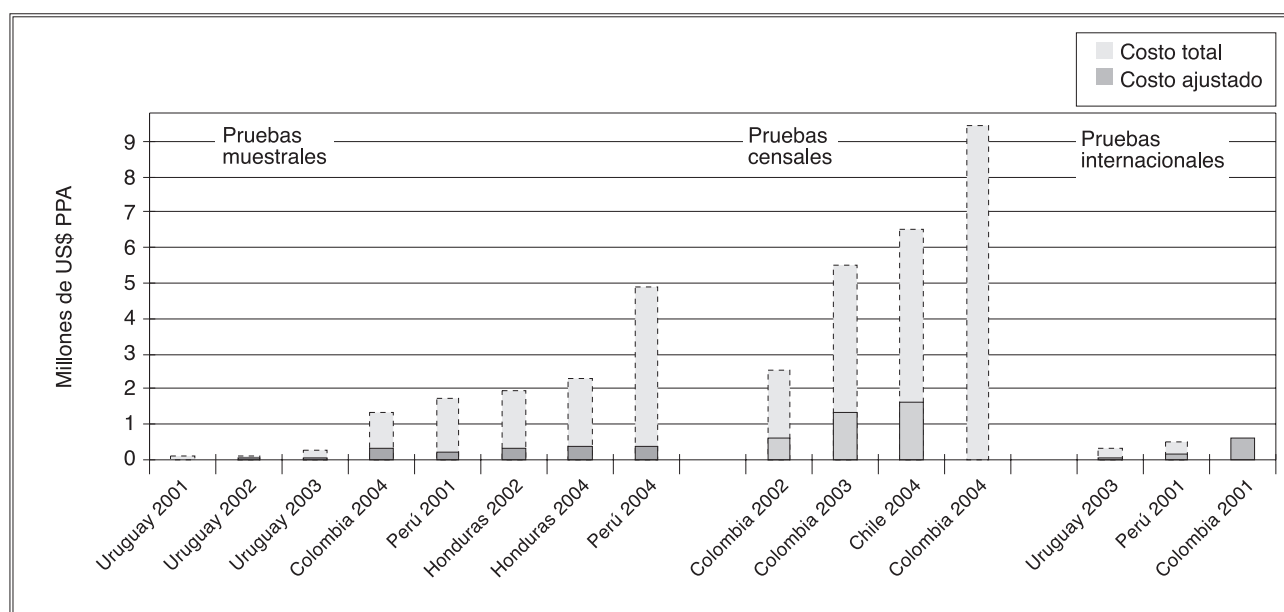
El costo por estudiante evaluado en muestras y en aplicaciones censales nacionales también varía significativamente, entre \$ 2 en Colombia y \$ 70 en Perú. Sin embargo, así como los costos totales, los costos relativos varían cuando el número de grados y materias evaluados son considerados, tal como se muestra en el gráfico 3.

Así pues, las explicaciones posibles de las diferencias de costos totales y unitarios de un país a otro podrían ser las siguientes:

- Evaluar múltiples materias y grados es más costoso que evaluar una o dos materias y un solo grado. La tabla 14 muestra un estimado de costos en base a una materia por grado. Con este enfoque, el costo estimado de evaluación en el Perú desciende significativamente en comparación con los otros países estudiados como se mostró en el gráfico 3. Por ejemplo, el Perú evaluó cuatro grados y tres materias en el 2004, es decir un total de 12 pruebas, comparadas con cuatro o seis en la mayoría de los otros países. Similarmente, en el 2001 Uruguay evaluó cuatro materias y tres grados, es decir un total de 12 pruebas (ver tabla 14).
- Países con poblaciones y territorios más grandes son más costosos de evaluar que países pequeños y compactos.
- Las preguntas abiertas son más costosas de calificar que las de opción múltiple. El que Colombia solo use preguntas de opción múltiple es un factor que explica sus costos relativamente más bajos, mientras que la decisión peruana de explorar los procesos por los cuales los estudiantes llegan a sus respuestas a los ítemes, usando cerca de 50% de preguntas abiertas, ciertamente contribuye a los mayores costos de este país.
- Los países con recursos humanos limitados pueden necesitar utilizar y pagar más a consultores locales y extranjeros, en comparación a países con mayor capacidad de recursos humanos.
- Aplicaciones piloto, controles de seguridad, así como análisis y difusión más cuidadosos y amplios incrementan los costos.
- La automatización de procesos y una administración eficiente pueden reducir los costos, según reportaron autoridades colombianas.
- El costo por estudiante de las evaluaciones de muestras pequeñas es mucho más alto que el costo de muestras más amplias o de todos los estudiantes.

GRÁFICO 2

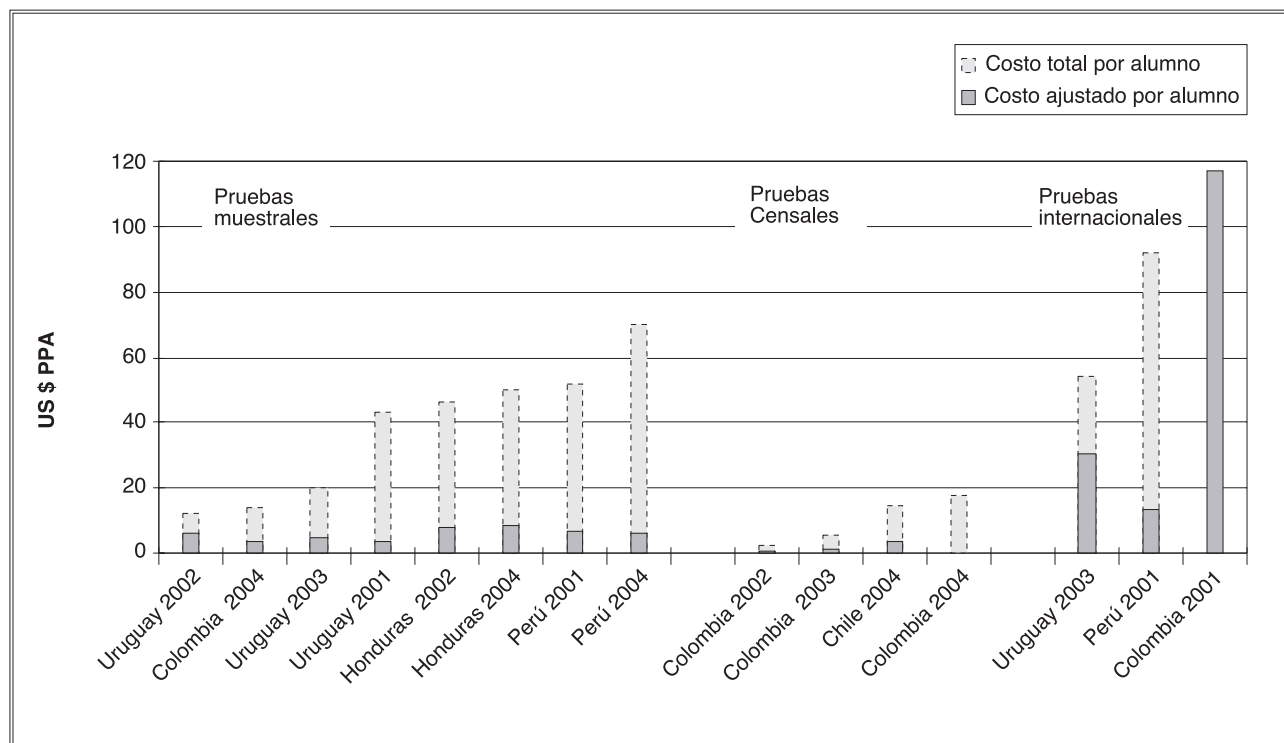
COSTOS TOTALES Y COSTOS AJUSTADOS



Nota: El costo ajustado considera el número de grados y materias evaluados se obtuvo dividiendo el costo total entre el número de grados y materias evaluados.

GRÁFICO 3

COSTOS TOTALES Y COSTOS AJUSTADOS POR ALUMNO EVALUADO



Nota: El costo ajustado por alumno se obtuvo dividiendo el costo por alumno entre el número de grados y materias evaluados.

TABLA 14

COSTOS DE EVALUACIÓN POR ÁREA DE ESTUDIO Y GRADO EVALUADOS

País, año de evaluación y tipo de evaluación	Costo unitario total (US\$)	Número de grados evaluados	Número de áreas de estudio	Número total de áreas de estudio evaluados	Costo unitario por grado y áreas de y áreas de estudio evaluados (US\$)
Chile 2004 C	15	1	4	4	3,75
Colombia 2001 I	117	1	1	1	117,00
Colombia 2002 C	2,5	2	2	4	0,63
Colombia 2003 C	5,3	2	2	4	1,25
Colombia 2004 M	14	2	2	4	3,50
Honduras 2002 M	46	2	3	6	7,70
Honduras 2004 M	50	2	3	6	8,30
Perú 2001 I	92	1	3	3	30,70
Perú 2001 M	52	3	2	6	8,67
Perú 2004 M	70	4	3	12	5,80
Uruguay 2003 I	54	1	4	4	13,50
Uruguay 2001 M	43	3	4	12	3,60
Uruguay 2002 M	12	1	2	2	6,00
Uruguay 2003 M	20	1	4	4	5,00

Nota: C = censal; M = muestral; I = internacional

- El apoyo de los organismos internacionales que efectúan préstamos o donaciones puede incrementar los costos debido a las regulaciones de adquisiciones como también simplemente por la disponibilidad de esos fondos.
- Hay incertidumbre sobre cómo dar cuenta de los costos indirectos de los organismos públicos que implementan el programa de evaluación de un país. En este informe no se incluyeron de manera completa y consistente esos costos indirectos en las tablas. Si se los incluyera, los costos podrían incrementarse hasta 60%, como en el caso de uno de los países estudiados. Chile calculó sus costos indirectos, pero no pudo distribuirlos entre las diferentes pruebas que estaba implementando al mismo tiempo. Colombia no estimó costos indirectos que probablemente son significativos. Perú estimó sus costos con y sin costos indirectos, que eran significativos. Honduras no tuvo costos indirectos gubernamentales, pues su unidad de evaluación estaba loca-

lizada en una universidad bajo contrato con el gobierno. Uruguay no estimó sus costos indirectos.

- Las escalas salariales pueden variar de un país a otro, incluso luego de ajustar en función a la paridad de poder adquisitivo (PPA).
- Las cuestiones relacionadas con la disponibilidad de personal pueden tener impactos financieros. La mayoría de los países de la región tiene muy reducidos cuadros de expertos en evaluación, que suelen estar sobrecomprometidos. Esto eleva los costos por la necesidad de contratar consultores externos caros, y puede crear problemas en la calidad técnica y en la difusión. A la larga, entrenar más expertos locales no solo mejorará la calidad técnica sino que bajará los costos.

En suma, rastrear los costos de las evaluaciones es complejo dadas las incertidumbres en la definición de lo que se considera costos indirectos y/o difusión, en la consideración de costos ocultos y la dificultad de hacer los costos comparables de un país a otro.

IV. COSTOS Y TOMA DE DECISIONES SOBRE LAS EVALUACIONES

Evaluar –al menos en base a una muestra– es un elemento fundamental en cualquier sistema educativo moderno, puesto que es un paso necesario en el proceso de diseñar, implementar y evaluar los esfuerzos de reforma¹⁰. Más allá de esta generalización, la toma de decisiones sobre los tipos y el alcance de las pruebas depende de una amplia gama de factores, empezando por la definición de objetivos viables de aprendizaje y otros, y luego pasando a consideraciones de capacidad técnica y administrativa. Los costos juegan un papel importante, pero no determinante, en estas decisiones. Cada país tiene un conjunto diferente de condiciones, y los encargados de la toma de decisiones y los técnicos precisan hacer sus propios cálculos de costo-beneficio sobre la amplitud y la profundidad de las pruebas, basados en esos objetivos. A continuación, presentamos algunas sugerencias a tomar en cuenta cuando los países ponderan qué tipo de evaluaciones implementar.

LA DIFICULTAD DE MEDIR RESULTADOS EN EDUCACIÓN

Los países deben reconocer que medir resultados en educación mediante evaluaciones no es un asunto simple. Un estudio reciente (Crone 2004) intenta resumir algunos problemas vinculados a la evaluación. En primer lugar, una prueba debe reflejar con precisión lo que se propone enseñar. Esto exige cuidado en el desarrollo de ítemes, pilotaje y análisis. En segundo lugar, un puntaje individual variará cada vez que un estudiante toma una prueba, no solo debido a los ítemes particulares incluidos en la prueba, sino también por la condición física y emocional del estudiante. Tercero, es difícil determinar a través de una evaluación si un maestro, una escuela o un sistema determinado es más efectivo que otro. El problema fundamental es que la materia prima, el estudiante que ingresa al sistema, no es homogéneo, pues los niños y jóvenes tienen distintos antecedentes, capacidades y experiencias. Por ello, un maestro cuyos estudiantes sacan puntajes promedio en una prueba estandarizada, pero cuyos padres son analfabe-

tos, podría considerarse mucho más efectivo que aquel cuyos estudiantes obtienen puntajes en el 25% superior, pero cuyos padres son ricos y con educación superior. También es difícil medir el valor agregado del aprendizaje de un año a otro, a menos que haya un número único de identificación de estudiantes. Algunos incrementos en los puntajes de las pruebas no reflejan incrementos de aprendizaje *per se*, sino más bien reflejan el incremento de experiencia y comodidad con el propio proceso de la toma de evaluaciones. Por último, los estados o los países pueden bajar sus estándares. En los Estados Unidos una serie de estados han reportado mejoras en los puntajes de las evaluaciones estatales. Al mismo tiempo, la Evaluación Nacional de Progreso Educativo (NAEP) ha mostrado poco o ningún progreso en esos mismos estados. Existe la sospecha de que estos estados están bajando sus estándares de evaluación para poder reportar incrementos de aprendizaje (Dillon 2005).

EQUILIBRANDO LO DESEABLE Y LO POSIBLE (*TRADE-OFFS*) EN TÉRMINOS DE METAS, COMPLEJIDAD, CAPACIDAD Y COSTOS DE LAS EVALUACIONES

Como se señaló antes, la toma de decisiones sobre las evaluaciones empieza con la identificación y selección de metas de aprendizaje y otros objetivos relevantes, la identificación de los actores interesados claves y la búsqueda de su interés y apoyo, y el reconocimiento de limitaciones reales en términos de recursos humanos y costos. Luego de considerar todas estas cuestiones, los encargados de la toma de decisiones necesitarán hacer y aceptar un balance de ventajas y desventajas relativas (*trade-offs*) sobre una amplia variedad de parámetros.

La regla de oro en este proceso es nunca escatimar en calidad técnica o en difusión, dado que una evaluación técnicamente pobre o aquella que no es usada o difundida es una real e inaceptable pérdida de dinero. Los objetivos de evaluación deben ser simples y claros.

En caso de incertidumbre, puede resultar mejor evaluar pocos grados y materias y concentrarse en la difusión. Evaluar muchos grados y materias a la vez, si bien ahorra en personal y costos de transporte y materiales, puede sobrepasar la capacidad del organismo evaluador para analizar, reportar y –lo más importante– difundir los resultados con efectividad; este quizás sea el

¹⁰ Uruguay es el mejor ejemplo en la región de un país que ha usado programas de evaluación como una línea de base para medir si una reforma importante ha tenido impacto en el aprendizaje con resultados positivos.

caso peruano. Pero, de otro lado, los maestros y estudiantes pueden considerar que aquellas materias que no fueron tema de evaluación carecen de importancia.

Otro tema crítico con el que hay que lidiar es si tomar muestras o aplicar pruebas de forma censal. Como lo ilustra la tabla 1, esta decisión debe tomarse en base a una clara definición de objetivos y de la forma en que se usarán los resultados. El tamaño de la muestra también tiene un fuerte impacto en el costo por estudiante evaluado. En Colombia, por ejemplo, una muestra de 5,000, para participar en evaluaciones internacionales, costó \$ 117 por estudiante. En una muestra mayor (96,000 estudiantes), el costo por estudiante evaluado fue \$ 14, y en pruebas censales está entre \$ 2 y \$ 5.

Debe tomarse en cuenta los costos fijos y los costos variables. Los costos fijos de una evaluación son la preparación y la validación de ítemes, incluida la aplicación de pruebas piloto. Los costos de preparación de ítemes se pueden reducir si solo una porción pequeña de estos se pone a disposición del público, lo cual permite reutilizar los ítemes ya validados. Los rubros de costo variable son la aplicación y la difusión de resultados. El análisis de resultados es una combinación de costos fijos y variables. Computadoras de gran poder permiten hoy en día analizar grandes bases de datos a costos adicionales pequeños, pero corregir ítemes abiertos es un costo variable. Corregir preguntas abiertas también puede demandar tiempo y costos, aunque suelen ser mejores medidas de capacidades de razonamiento más complejas. Cuando Colombia realizó una evaluación censal, sus costos fijos de preparación de las pruebas alcanzaron solo 6.6% de los costos totales. Cuando realizó una muestra, sus costos fijos fueron 10.6% del costo total. En el estudio PIRLS de menos de 6,000 estudiantes, los costos de preparación de las pruebas fueron 25.4% de los costos totales. Las muestras pequeñas pueden ser costo-efectivas, pero, al mismo tiempo, muestras grandes pueden ampliarse hasta una evaluación censal a bajo costo marginal, puesto que los costos fijos de desarrollo de ítemes y pruebas piloto pueden ser amortizados sobre una población más grande.

Si tomamos en cuenta la capacidad instalada actual, evaluar a todos los estudiantes de cada grado cada año, tal como se hace en muchos

estados de los Estados Unidos, parece no ser una buena idea para los países de América Latina. Chile tiene probablemente el programa de pruebas censales más amplio, pero hasta hace poco habían estado evaluando no más de un grado al año, entre 1º y 8º. Implementar un programa completo de evaluaciones anuales en Chile incrementaría los gastos cerca de ocho veces, lo cual llevaría a que los costos sean más de 1% de los gastos educativos nacionales, aún suponiendo algunas economías de escala. En Perú y Honduras el costo proporcional sería aún más alto. En comparación, el costo en Estados Unidos es, en promedio, equivalente al 0.25% del gasto total anual por estudiante. Este bajo ratio en Estados Unidos es, claramente, resultado de sus gastos mucho más altos en la enseñanza y los maestros.

Si los costos educativos se incrementaran en 1% por las evaluaciones, y ello pudiera dar lugar a grandes incrementos en los logros de los estudiantes, por cierto que valdría la pena. Pero es probable que la evaluación masiva distraiga de la tarea más importante de difundir los resultados de la prueba y diseñar la formación de maestros y otras intervenciones basadas en los resultados de la evaluación, sin lo cual no habría ninguna mejora, además de exigir montos adicionales considerables del gasto público. Evaluar solo algunos grados y materias revelará avances y desafíos para las escuelas, comunidades, provincias y el país como un todo, y ayudará a los padres a ubicar a sus hijos en un contexto más amplio. Es decir, evaluar uno o dos grados por año logra casi todas las metas de la evaluación educativa salvo la retroalimentación a todos los padres. Evaluar a todos los estudiantes de todos los grados todos los años incrementa los aspectos “negativos” de este proceso, como son el disminuir el tiempo dedicado a la enseñanza, descuidar materias no evaluadas (como historia, artes y ciencias) y desalentar la innovación a nivel de aula.

ENFATIZANDO LA DIFUSIÓN DE LOS RESULTADOS

Sin una difusión de resultados adecuada, la evaluación no tendría impacto alguno sobre el aprendizaje o rendimiento, puesto que los resultados no serán utilizados por los maestros, los padres, los directores de escuela ni los tomadores de decisiones. Debido a ello, no debería escatimarse nunca en los gastos en la difusión y deberá contarse con un plan de difusión antes

de poner en marcha cualquier proceso de evaluación. En el caso latinoamericano se observa que la difusión no es costosa comparada con el resto del proceso de evaluación. Los costos reportados de la difusión, usualmente para la publicación de resultados, fluctúa entre 5% a 15% de los costos totales de las evaluaciones. Un costeo exhaustivo de la difusión debería incluir también el análisis que deberían realizar los tomadores de decisiones y los cambios en la formación docente, el currículo y los libros de texto. Para una discusión detallada sobre el estado actual de la difusión de los resultados de las evaluaciones en América Latina véase Ravela 2002 y Cueto 2005.

EL IMPACTO POTENCIAL DE LAS PRUEBAS DE ALTAS CONSECUENCIAS

Las pruebas de altas consecuencias a nivel de la secundaria, o exámenes de fin de curso basados en el currículo, determinan si un estudiante se graduará de un nivel particular de educación y puede tener impacto en sus opciones de educación superior. Esto es común en países tan diversos como Gran Bretaña, Francia, Dinamarca, Japón, Corea del Sur, Holanda y Alemania. Son distintos a los exámenes de ingreso universitario porque son universales, necesarios para graduarse y se basan en el currículo. Hay evidencia internacional creciente que las pruebas de altas consecuencias diseñadas para certificar a los estudiantes que han culminado satisfactoriamente sus estudios secundarios pueden por sí mismas incrementar el nivel de logro de aprendizaje. Tres estudios recientes (Bishop 1997, 2003 y Woessman 2000) concluyeron que las pruebas de altas consecuencias incrementaban el aprendizaje a nivel de la secundaria. De los 40 países que participaron en la prueba internacional TIMSS, los países con pruebas nacionales de altas consecuencias generalmente obtuvieron puntajes más altos que aquellos que no las tenían¹¹. En Estados Unidos, estados con ese tipo de pruebas, como Nueva York (el examen Regents al final de la secundaria), tienen mejores resultados que los estados sin exámenes de altas consecuencias. Estos resultados son independientes de los gastos por estudiante.

Estas conclusiones son advertidas cada vez más en Estados Unidos. Por ejemplo, el Scholastic Aptitude Test (SAT) en Estados Unidos está basado cada vez más en logros reales antes que en destrezas de aprendizaje “general”, que es una manera de incrementar las “consecuencias” del aprendizaje de la secundaria. Además el Advanced Placement Exam del Servicio de Evaluación Educativa (ETS) de Estados Unidos, que también es una prueba de altas consecuencias para los estudiantes de secundaria que buscan entrar a los *colleges* y universidades de élite, se está volviendo cada vez más generalizado en Estados Unidos.

Todos los países caribeños angloparlantes tienen exámenes de altas consecuencias, no así la mayoría de los países latinoamericanos de la región. Las excepciones son Costa Rica y República Dominicana, donde una prueba rendida al final de la secundaria pesa por lo menos 25% en la nota final del estudiante, y El Salvador, donde la Prueba de Aptitudes para Egreso de Secundaria (PAES) determina 20% del calificativo final. Todavía no hay evidencias de que los estudiantes de estos países se desempeñen mejor que lo esperado debido a estos exámenes. Chile ha empezado a considerar evaluaciones de altas consecuencias al final de la educación secundaria y el Consejo Nacional del Perú ha propuesto que se considere seriamente esta opción. Este informe no aporta datos sobre los costos de estas pruebas, con excepción de ICFES. Es posible que las pruebas sean relativamente caras, pero la recompensa en términos de logros de aprendizaje puede ser significativa.

Las evaluaciones de altas consecuencias tienen riesgos obvios. Se debe cuidar que las pruebas estén diseñadas para enfatizar aprendizajes de alto nivel, antes que la memorización. También se debe asegurar que los estudiantes no abandonen los estudios antes del examen de altas consecuencias. Esto ocurrió hace unos años en Texas, aunque se reporta que las deserciones escolares se han estabilizado (Carnoy y Loeb 2003). También existen reportes anecdóticos de deserciones previas a las evaluaciones en el Caribe angloparlante.

Una variación de las evaluaciones de altas consecuencias usada en Estados Unidos son los exámenes de “competencia mínima”, en los cuales todos los estudiantes deben alcanzar un

11 El estudio de Bishop también concluyó que otra razón del bajo desempeño de los estudiantes de EE.UU. es que sus maestros están gravemente subremunerados y sobreexigidos en comparación a los de otros países.

“piso” de aprendizaje para poder graduarse. La evidencia disponible hasta ahora (ver Bishop 1997) es que los exámenes de competencia mínima no tienen un impacto observable sobre el aprendizaje y pueden producir un incremento en la tasa de deserción escolar.

EL VALOR DE PARTICIPAR EN PROGRAMAS INTERNACIONALES DE EVALUACIÓN

La participación en pruebas internacionales puede generar un debate nacional y esfuerzos de reforma que pueden llevar a mejorar el aprendizaje y representan un espacio valioso para el desarrollo de capacidades nacionales en evaluación. Los puntajes bajos en las evaluaciones han conducido a tales debates en todo el mundo y a la generación de una amplia variedad de programas para el mejoramiento del aprendizaje en los países participantes. Por ejemplo, Alemania alcanzó puntajes mucho más bajos que los esperados en el examen PISA y ahora está repensando sus políticas de canalización temprana de muchos estudiantes a programas de bajas expectativas (Ammermueller 2004). Algunos lugares (por ejemplo Corea del Sur, Hong Kong, Estados Unidos y Ontario, entre otros) han mejorado aprendizajes recientemente, según sus puntajes en la evaluación TIMSS de ciencias y matemáticas entre 1995 y 2003. Las razones de dichas mejoras requieren un análisis que trasciende el foco del presente estudio. En América Latina, se han suscitado debates nacionales similares alrededor de los resultados pobres de Chile, México, Perú y Brasil tanto en TIMSS como en PISA, aunque aún no se observan resultados positivos de dichas discusiones. Es posible que los estudiantes latinoamericanos tengan resultados bajos debido a la falta de oportunidades de aprendizaje, es decir, porque aquello que se mide en las pruebas no está siendo enseñado en las aulas. Por ejemplo, muchas de las pruebas internacionales de lenguaje requieren que el estudiante escriba ensayos o, para matemáticas, que solucionen problemas, habilidades que son poco desarrolladas en la típica aula latinoamericana.

Si bien la participación en programas internacionales de evaluación no es una carga financiera significativa, y es muy valorada como terreno de entrenamiento por casi todos los participantes de América Latina, hay otros problemas que con frecuencia dificultan la participación de algunos países y que deben tomar en cuenta los encar-

gados de la toma de decisiones. Por ejemplo, las reuniones de PISA y PIRLS pueden celebrarse en cualquier lugar del mundo. Viajes largos, digamos al Lejano Oriente, a menudo tienen que recibir el visto bueno de funcionarios gubernamentales del más alto nivel, sobre todo en países pequeños, y puede ser difícil cuando los países están implementando medidas de austeridad fiscal. Los encargados de tomar decisiones educativas tienen que convencer a los funcionarios de presupuesto sobre la importancia de estos programas internacionales.

Además, las exigencias planteadas a grupos pequeños de expertos nacionales calificados, usualmente involucrados simultáneamente en esfuerzos de evaluación nacional, pueden ser muy altas. La solución a largo plazo es capacitar a más especialistas en el área de evaluación.

Para algunos países parece ser más atractivo participar en las pruebas regionales de UNESCO, sobre todo debido a la posibilidad de un involucramiento cercano, aprender haciendo y menores costos. Las pruebas de UNESCO podrían ayudar a alcanzar muchas de las metas de las pruebas internacionales si porciones de ellas se hicieran equivalentes a PISA, PIRLS, etc.

NECESIDAD DE IDENTIFICAR “QUÉ FUNCIONA” EN LA EDUCACIÓN

Las evaluaciones estandarizadas no son caras si se las compara con los gastos totales en educación en América Latina. Una cuestión más fundamental es identificar qué es “lo que funciona” en educación. Desafortunadamente, los sistemas educativos alrededor del mundo resaltan por invertir muy poco en investigación y desarrollo. América Latina no es una excepción.

Como se señaló antes, la investigación sugiere que las pruebas de altas consecuencias pueden incrementar el aprendizaje. Hay evidencia sugere —aunque no concluyente— de que la retroalimentación de los resultados de las pruebas que conllevan recompensas/sanciones para los maestros y escuelas (es decir, responsabilización o *accountability*) en función a mejoras en el aprendizaje afectarán positivamente ese aprendizaje. En los Estados Unidos, la ley *No Child Left Behind* ordena acciones para mejorar los puntajes de las evaluaciones e incluye una variedad de recompensas y sanciones a las escuelas y comunidades por cambios en puntajes.

Un estudio reciente (Carnoy y Loeb 2003) concluyó que, si bien ello parece prometedor, aún es demasiado temprano para identificar con claridad si la responsabilización del maestro, escuela o comunidad en Estados Unidos ha tenido impacto sobre el aprendizaje. En la región, Chile ha identificado distritos de bajo desempeño (el programa de las 900 escuelas) para que reciban ayuda adicional. Anteriormente, ha habido mejoras en estas escuelas pero también hay evidencia de que algunas escuelas no deseaban “graduarse” a mejores niveles de desempeño para no perder esos recursos adicionales. El sistema chileno para estimular el desempeño docente (el SNED) es bien aceptado por los maestros, pero aún no se ha demostrado que tenga un impacto en el aprendizaje. Tampoco en México hay evidencias de que los incentivos a los maestros que obtienen mayores puntajes en las evaluaciones hayan mejorado los logros de aprendizaje (ver Mizala y Romaguera 2005 para un estudio de la experiencia de ambos países). De igual manera, la retroalimentación de los resultados de las pruebas para reformar el currículo, revisar los libros de texto y mejorar la capacitación del maestro debería tener un impacto sobre el aprendizaje, pero existe muy poca evidencia directa de ese impacto, en gran medida por lo complejo que es el proceso. En suma, queda mucho por hacer en cuanto a medir el impacto de los programas y políticas educativas sobre el aprendizaje.

BIBLIOGRAFÍA

- Ammermuller, Andreas. 2004. “PISA: What Makes the Difference? Explaining the Gap in PISA Test Scores between Finland and Germany”. Social Science Research Network, Center for European Economic Research, Munich 2004.
- Bishop, John. 2003. “La educación secundaria en los Estados Unidos. ¿Qué pueden aprender otros de nuestros errores?”, en Rama (ed.) *Alternativas de reforma de la educación secundaria*, ed. German Rama, IDB, Washington DC 2003.
- Bishop, John. 1997. “The Effect of National Standards and Curriculum-Based External Exams on Student Achievement”, *American Economic Review*, mayo 1997.
- Bruneforth, Michael, Albert Motivans y Yanhong Zhang. 2004. “Investing in the Future: Financing the Expansion of Educational Opportunity in Latin America and the Caribbean”, UIS, Unesco. Montreal.
- California Legislative Analyst's Office, “Analysis of the 2003-2004 Budget Bill, Education Table of Contents”, 2003 Sacramento.
- Carnoy, Martin, y S. Loeb. 2003. “Does Accountability Affect Student Outcomes? A Cross State Analysis”, *Education Evaluation and Policy Analysis 2003*. American Education Research Association.
- Carnoy. 2004. ¿Tiene efecto la responsabilidad externa en los resultados educacionales de los alumnos? Un análisis entre los estados de EE.UU. Documento de Trabajo N° 26. Santiago, PREAL.
- Cohen, Joel, y David Bloom. 2005. “Cultivating Minds,” *Finance and Development*, junio 2005, IMF, Washington DC.
- Connecticut State Department of Education. 2005. “Cost of Implementing the Federal No Child Left Behind Act in Connecticut”, marzo 2005, Hartford, Connecticut.
- Crone, Theodore. 2004. “What Test Scores Can and Cannot Tell Us about the Quality of our Schools”, *Federal Reserve Bank of Philadelphia Business Review*, Q3, 2004. Filadelfia.
- Cueto, Santiago, ed., “Uso e impacto de la información educativa en América Latina”, PREAL 2005, Santiago, Chile.
- Chile. 2005. “SIMCE 2004 Octavo básico. Resultados nacionales”, www.simce.cl
- Dillon, Sam. 2005. “Students Ace State Tests, but Earn D's from US”, *New York Times*, Nov 26, 2005.
- Ferrer, Guillermo. 2004. “Estado de situación de los sistemas nacionales de evaluación de logros de aprendizaje en América Latina”. PREAL, Santiago, Chile.

Mizala, Alejandra, and Pilar Romaguera. 2003. "Rendimiento escolar y premios por desempeño: la experiencia latinoamericana," PREAL, Santiago, Chile.

Peyser, James y Robert Costrell. 2004. "Exploring the Costs of Accountability," *Education Next*, Hoover Institution, Stanford, California.

Ravela, Pedro (ed). 2001. Los próximos pasos: ¿Cómo avanzar en la evaluación de aprendizajes en América Latina? Santiago, Chile, Documento de Trabajo de PREAL N° 20. Agosto 2001.

Ravela, Pedro. 2002. "¿Cómo presentan sus resultados los sistemas nacionales de evaluación educativa en América Latina?", Documento de Trabajo de PREAL N° 22, Santiago, Chile.

Schiefelbein, Ernesto, Laurence Wolff y Paulina Schiefelbein. 2000. "The Opinion of Experts as an Instrument to Evaluate Investment in Primary Education", *Revista de la CEPAL*, diciembre del 2000, Santiago, Chile.

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

Woessmann, Ludger. 2000. "Schooling Resources, Educational Institutions, and

Student Performance: the International Evidence," mayo 2000, Kiel Working Paper N° 983, Kiel Institute of World Economics. Alemania.

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

SITIOS WEB

Chile: Sistema de Medición de la Calidad de la Educación (SIMCE) www.simce.cl

Colombia: Instituto Colombiano de Fomento de la Educación Superior www.icfes.gov.co

Honduras; Unidad de Medición de la Calidad Educativa www.se.gob.hn

Perú: Unidad de Medición de la Calidad Educativa www.minedu.gob.pe/umc

Uruguay: Administración Nacional de la Educación Pública www.anep.edu.uy
www.pisa.uruguay@adinet.com.uy

UNESCO/Orealc: Laboratorio Latinoamericano de Medición de la Calidad de la Educación <http://llece.unesco.cl/esp>

Anexo

RESUMEN DE SISTEMAS NACIONALES DE EVALUACIONES DE APRENDIZAJES

Los siguientes textos y tablas fueron extractados de Ferrer (2006) e incluyen información básica sobre las características de los sistemas nacionales de evaluación de la educación de los países latinoamericanos incluidos aquí –Chile, Colombia, Honduras, Perú y Uruguay– que se deberían tener en cuenta cuando se comparan los costos.

CHILE

Las pruebas son censales en todos los grados de aplicación. Históricamente han sido diseñadas con referencia a normas, aunque recientemente se ha redefinido el modelo para poder reportar los resultados con referencia a criterios. Los instrumentos incluyen preguntas cerradas de opción múltiple, así como preguntas para el desarrollo de respuestas abiertas. Todos los operativos son acompañados con cuestionarios de contexto para el estudio posterior de factores

escolares y extraescolares asociados al rendimiento académico. Complementariamente, un equipo técnico especial realiza un seguimiento cuanti-cualitativo de implementación curricular a nivel nacional mediante encuestas y observaciones a docentes y directores de escuelas. Ese seguimiento permite indagar más profundamente sobre la incidencia que las diferentes fuentes de prescripción curricular utilizadas en la escuela (marcos curriculares, programas de estudio, libros de texto, etc.) tienen sobre el logro académico de los alumnos.

CUADRO A-1

CHILE- RESUMEN DEL SISTEMA NACIONAL DE EVALUACIÓN

Años de aplicación	Grados	Áreas disciplinares	Tipo de Estudio	Publicidad de resultados para conocimiento de	Uso de resultados para
Pruebas nacionales: Prueba de Evaluación de Rendimiento Escolar					
1982-1984	4,8	· Lenguaje · Matemáticas · Ciencias · Naturales · Ciencias sociales	· Censo nacional · Muestra nacional	· Gobierno · Usuarios · Público	---
Prueba nacional, aplicado por SIMCE					
1988	4	· Lenguaje · Matemáticas · Ciencias naturales · Ciencias sociales · Aspectos Actitudinales	· Censo Nacional · Muestra experimental	· Gobierno · Usuarios · Público	---
1989	8				
1990	4				
1991	8				· Capacitación docente · Desarrollo curricular · Focalización de apoyos a estudiantes y centros educativos
1992	4				
1993	2M	· Aspectos actitu- dinales			
1994	4,8	· Lenguaje · Matemáticas · Ciencias naturales · Ciencias Sociales · Aspectos actitudinales			
1995	2M	· Lenguaje · Matemáticas · Aspectos actitudinales			

	8	· Lenguaje · Matemáticas · Ciencias naturales · Ciencias sociales			
1996	4	· Lenguaje · Matemáticas			
1997	8	· Ciencias naturales · Ciencias sociales · Aspectos actitudinales			
1998	2M	· Lenguaje · Matemáticas	· Muestra nacional · Muestra experimental	· Gobierno · Usuarios · Público	
1999	4	· Lenguaje · Matemáticas	· Censo Nacional · Muestra experimental		
2000	8	· Ciencias naturales · Ciencias sociales			
2001	2M	· Lenguaje · Matemáticas			
2002	4	· Lenguaje · Matemáticas · Ciencias naturales · Ciencias sociales			
2003	2M	· Lenguaje · Matemáticas			
2004	8	· Lenguaje · Matemáticas · Ciencias naturales · Ciencias sociales	· Censo nacional		· Capacitación docente · Desarrollo curricular · Focalización de apoyos a estudiantes y centros educativos · Incentivos docentes
Pruebas Internacionales					
Estudio de seis materias (1970-71) LLECE (1997) IEA-Civic Education (2000) Pisa Plus (2001)					

COLOMBIA

Las pruebas SABER son muestrales, están referidas a criterios y evalúan el rendimiento académico de acuerdo a una escala predefinida de tres niveles de desempeño de complejidad creciente para cada asignatura y año evaluados. Las pruebas se complementan con cuestionarios de contexto que permiten el análisis de factores escolares y extraescolares asociados al rendimiento académico (características de escuela, maestro, alumno y fami-

lia). Los Exámenes de Estado se aplican censalmente a todos los aspirantes a la educación superior. Las pruebas miden el alcance de conocimientos en las áreas curriculares de formación general, y también permiten la evaluación de conocimientos en profundidad para algunas disciplinas, según los requisitos de ingreso de los diferentes programas de educación superior. Se evalúan también conocimientos en lenguas extranjeras y formación interdisciplinaria (específicamente, medio ambiente y violencia y sociedad).

CUADRO A-2

COLOMBIA- RESUMEN DEL SISTEMA NACIONAL DE EVALUACIÓN

Años de aplicación	Grados	Áreas disciplinares	Tipo de Estudio	Publicidad de resultados para conocimiento de	Uso de resultados para
Pruebas nacionales: aplicado por ICFES/ Ministerio de Educación					
1991-1994	3, 5, 7, 9	· Lenguaje · Matemáticas	· Muestra nacional · Muestra regional	· Gobierno · Usuarios · Público	-
Pruebas nacionales: SABER					
1997-1998	3, 5, 7, 9	· Lenguaje · Matemáticas	Muestra nacional	· Gobierno · Usuarios · Público	-
1998-1999	7, 9	· Lenguaje · Matemáticas			
2002-2003	5, 9*	· Ciencias naturales	Censo nacional	-	
Pruebas nacionales: ICFES Exámenes de Estado					
1980-2005	11	· Lenguaje extranjero · Lenguaje · Matemáticas · Ciencias naturales · Ciencias sociales	Censo nacional	· Gobierno · Usuarios	· Desarrollo curricular · Selección educación superior
Pruebas internacionales					
· TIMSS (1995) · LLECE (1997) · IEA-Civic Education (2000) · PIRLS (2001)					

* En algunas entidades territoriales se realizó, adicionalmente, la evaluación de los grados 3 y 7.

HONDURAS

Las pruebas aplicadas por la UMCE son muestrales a nivel nacional, referidas a criterios y basadas principalmente en un modelo de respuesta de opción múltiple. En el área de lenguaje, además, se evalúa redacción escrita y lectura oral. Las pruebas son complementadas con cuestionarios a alumnos, padres de familia, docentes y directores de centros, para el estudio posterior de factores asociados al rendimiento. Se estudian variables de NES y otras variables intraescolares (incluyendo un modelo de medi-

ción de eficacia escolar). Los equipos técnicos han encontrado dificultades para establecer correlaciones confiables entre esas variables y el rendimiento académico, aunque la incorporación de un modelo de análisis jerárquico lineal (HLM) a partir de 2002 ha permitido aumentar la confiabilidad y validez de las correlaciones. Además del programa regular de evaluación en los grados y áreas mencionadas, la UMCE presta sus servicios para la medición de rendimiento académico como indicador de impacto de algunos programas específicos financiados por agencias de cooperación internacional.

CUADRO A-3

HONDURAS- RESUMEN DEL SISTEMA NACIONAL DE EVALUACIÓN

Años de aplicación	Grados	Áreas disciplinares	Tipo de Estudio	Publicidad de resultados para conocimiento de	Uso de resultados para
Pruebas nacionales: Aplicado por la Secretaría de Educación*					
1990-1994	1, 2, 3, 4, 5	· Lenguaje · Matemáticas · Ciencias naturales · Ciencias sociales	· Muestra nacional (10 municipios)	· Secretaría de Educación · USAID	· Evaluar proyectos apoyados por USAID
Pruebas nacionales, aplicado por UMCE					
1997	3, 6	· Lenguaje · Matemáticas	· Muestra nacional · Censo nacional	· Gobierno · Usuarios	· Capacitación docente · Focalización de apoyos a estudiantes y centros educativos
1998	2, 3, 4, 6		· Muestra nacional		
1999	2, 3, 4, 5				
2000	3, 6				
2002-2004	3, 6				
Pruebas internacionales					
LLECE (1997)					

* Prueba fue parte de componente V del Proyecto PEEP (Eficiencia de la Educación Primaria)

PERÚ

Las primeras aplicaciones se realizaron con pruebas referidas a normas, pero actualmente se aplican instrumentos con referencia a criterios. Las pruebas son muestrales, incluyen ítems de opción múltiple y de respuestas abiertas, y han sido aplicadas en español, quechua y aymara. Junto con las pruebas se aplican cues-

tionarios de contexto para el estudio posterior de factores asociados al rendimiento. Se estudian variables escolares (Ej. Insumos pedagógicos, características de escuelas y profesores, actitudes hacia las asignaturas y hacia las lenguas originarias, tareas en el hogar), y extraescolares (principalmente género, NES, lengua materna, localización urbana o rural, administración pública o privada, y región geográfica).

CUADRO A-4

PERÚ- RESUMEN DEL SISTEMA NACIONAL DE EVALUACIÓN

Años de aplicación	Grados	Áreas disciplinares	Tipo de Estudio	Publicidad de resultados para conocimiento de	Uso de resultados para
Pruebas nacionales: Aplicado por UMC					
1996	4	· Lenguaje · Matemáticas	· Muestra nacional	-	-
1998	4, 6, 4M, 5M	· Lenguaje · Matemáticas · Ciencias naturales* · Ciencias sociales*		· Gobierno · Usuarios · Público	
2001	4, 6, 4M	· Lenguaje · Matemáticas			
2004	2, 6, 3M, 5M	· Lenguaje · Matemáticas · Ciudadanía			
Pruebas internacionales					
· LLECE (1997) · PISA Plus (2001)					

* Aplicado solamente a grados 4 y 6.

URUGUAY

Se han realizado operativos censales y muestrales. Las pruebas incluyen ítemes de opción múltiple y también preguntas para el desarrollo de respuesta abierta. Junto a las pruebas se aplican cuestionarios de contexto socioeducativo a docentes, directores, alumnos y familias para el es-

tudio de factores asociados al rendimiento. Se analizan variables escolares (Ej. Infraestructura y equipamiento, dotación de recursos humanos, experiencia docente, gestión, concepciones pedagógicas), así como variables extraescolares (Ej. Condiciones de la vivienda, composición de la familia, bienes materiales y culturales, niveles educativos y ocupacionales de los padres).

CUADRO A-5

URUGUAY- RESUMEN DEL SISTEMA NACIONAL DE EVALUACIÓN

Años de aplicación	Grados	Áreas disciplinares	Tipo de Estudio	Publicidad de resultados para conocimiento de	Uso de resultados para
Pruebas nacionales: Aplicado por la Unidad de Medición de la Calidad de la Educación					
1996	6	· Lenguaje · Matemáticas · Aspectos actitudinales	Censo nacional	· Gobierno · Usuarios · Público	· Capacitación docente · Focalización de apoyos a estudiantes y centros educativos
1998	3	· Lenguaje · Matemáticas · Ciencias naturales · Ciencias sociales	· Muestra nacional · Autoaplicación universo escuelas	· Usuarios · Público	· Capacitación docente
1999	6	· Lenguaje · Matemáticas · Aspectos actitudinales	· Muestra nacional · Muestra experimental · Autoaplicación universo escuelas	· Gobierno · Usuarios · Público	
1999	3M	· Lenguaje · Matemáticas · Ciencias naturales · Ciencias sociales	· Censo nacional		
2001	4	· -	· -	-	· -
2001	Preescolar, 1, 2	· Desarrollo cognitivo y afectivo · Lenguaje · Matemáticas	· Muestra nacional	· Gobierno · Usuarios	· Capacitación docente
2002	6	· Lenguaje · Matemáticas	· Censo nacional · Muestra nacional	-	-
Pruebas internacionales					
PISA (2003)					



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

El Programa de Promoción de la Reforma Educativa en América Latina y el Caribe es un proyecto conjunto del Diálogo Interamericano, con sede en Washington, y la Corporación de Investigaciones para el Desarrollo, con sede en Santiago de Chile.

Desde su creación en 1995, el PREAL ha tenido como objetivo central contribuir a mejorar la calidad y equidad de la educación en la región mediante la promoción de debates informados sobre temas de política educacional y reforma educativa, la identificación y difusión de buenas prácticas y la evaluación y monitoreo del progreso educativo.

La ejecución de las actividades se realiza a través de Centros Asociados de Investigación y Políticas Públicas en diversos países de la región y comprenden la realización de estudios, la organización de debates y la promoción de diálogos públicos sobre temas de política educacional y reforma educativa.

Las actividades del PREAL son posibles gracias al apoyo de la United States Agency for International Development (USAID), el Banco Interamericano de Desarrollo (BID), la International Association for the Evaluation of Educational Achievement (IEA), GE Foundation, The Tinker Foundation, entre otros.



Inter-American Dialogue • 1211 Connecticut Ave. N.W. Suite 510
Washington, D.C. 20036 U.S.A. • Tel: (202) 822-9002
Fax: (202)822-9553 • E-mail: iad@thediologue.org
Internet: www.thediologue.org & www.preal.org

CINDE • Santa Magdalena 75, Piso 10 • Oficina 1002 • Providencia
Santiago, Chile • Tel: (56-2) 334-4302
Fax: (56-2) 334-4303 • E-mail: infopreal@preal.org
Internet: www.preal.org

