



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

N° 26

**Las pruebas internacionales
de aprendizaje en América
Latina y su impacto en la
calidad de la educación:
Criterios para guiar futuras
aplicaciones**

**J. Guillermo Ferrer
Patricia Arregui**

Abril 2003

El presente documento corresponde a un resumen del estudio del mismo nombre financiado con recursos del Consorcio de Investigación Económica y Social del Perú, y con el apoyo de PREAL. La versión completa se encuentra en <http://www.grade.org.pe/gtee-preal/docr/ExperLatino.pdf>

Las opiniones vertidas en este trabajo son de responsabilidad de los autores y no comprometen al PREAL, ni a las instituciones que lo patrocinan.

ÍNDICE

INTRODUCCIÓN	5
I. LAS PRUEBAS INTERNACIONALES DE APRENDIZAJE	5
I.1. Las reformas educativas y la evaluación del logro académico en los noventa	5
I.2. ¿En qué consisten las pruebas internacionales?	8
I.3. El debate acerca de la participación en pruebas internacionales	11
II. MARCO CONCEPTUAL Y CASOS DE ESTUDIO	13
III. RESULTADOS DEL TRABAJO DE CAMPO	14
III.1. Agencia: LLECE (Unesco-OREALC) - Primer Estudio Internacional Comparativo sobre Lenguaje, Matemáticas y Factores Asociados	14
III.2. Agencia: IEA - TIMSS, TIMSS-R y Educación Cívica	15
III.3. Agencia: OCDE - Prueba PISA	17
IV. RESUMEN CRÍTICO DE LOS RESULTADOS	17
V. RECOMENDACIONES PARA FUTURAS OPORTUNIDADES DE PARTICIPACIÓN EN PRUEBAS INTERNACIONALES	22
BIBLIOGRAFÍA	27
ENTIDADES ENTREVISTADAS	28
SIGLAS Y ABREVIACIONES	29

INTRODUCCIÓN

En los últimos años, diversos organismos internacionales y regionales han recomendado que los países latinoamericanos participen en pruebas internacionales de logro académico. Esa recomendación suele sustentarse en el argumento de que los datos comparativos sobre el rendimiento estudiantil son útiles para informar a los países sobre la calidad relativa de sus sistemas educativos y, consecuentemente, sobre su competitividad en el mercado global de bienes y servicios. Se argumenta también que las pruebas internacionales garantizan mayor calidad técnica y eficiencia que los sistemas nacionales de medición y que pueden contribuir significativamente al desarrollo de capacidades locales en el campo de la evaluación.

La utilidad efectiva de las comparaciones internacionales de logro académico para informar políticas educativas no es, sin embargo, algo sobre lo cual exista total acuerdo en la literatura actual. Tampoco se ha recogido ni sistematizado suficiente evidencia empírica sobre la difusión que los países han dado a la información recogida en las pruebas internacionales ni sobre la manera en que han utilizado esa información para promover el mejoramiento de sus políticas, planes, programas y prácticas educativas.

Ante esta falta de evidencia, y ante la posibilidad de que los países de la región comiencen a privilegiar la participación en pruebas internacionales sin consolidar el desarrollo de sus propios sistemas y capacidades de evaluación y medición, se consideró oportuno iniciar una indagación exploratoria sobre los riesgos y beneficios de esta participación y sobre las condiciones técnicas y políticas en que se han venido realizando.

Para ello se decidió estudiar la experiencia de algunos de los países que han participado en pruebas internacionales de logro académico: Argentina, Chile, Colombia, Cuba, Ecuador, Estados Unidos, México, Perú y Uruguay (que participará próximamente por primera vez en una prueba internacional). La inclusión de Estados Unidos se decidió porque su experiencia arroja cierta luz sobre la forma en que América Latina puede hacer mejor uso de sus oportunidades de participación en pruebas internacionales. Así, aunque esta muestra no se considera representativa de toda la región, permite identificar algunas tendencias y patrones comunes en los pro-

cesos de participación de América Latina en pruebas internacionales¹.

El propósito último de este estudio es el de contribuir a que dicha participación, de resultar recomendable o de alguna manera ineludible, pueda servir para el mejoramiento de la gestión y de los resultados de los procesos educativos. A la vez, confiamos en que las recomendaciones finales de este estudio servirán para minimizar los impactos negativos que dichas actividades pudieran eventualmente acarrear.

Antes de concluir esta introducción, deseamos agradecer a Gilbert Valverde, Fernando Reimers, Juan Fernando Vega, José Rodríguez, Leonor Cariola, Carlos Pardo, Silvia Montoya, Shane Hunt, Pedro Ravela y Héctor Fernández por sus valiosos comentarios a los primeros borradores de este informe.

I. LAS PRUEBAS INTERNACIONALES DE APRENDIZAJE

I.1. LAS REFORMAS EDUCATIVAS Y LA EVALUACIÓN DEL LOGRO ACADÉMICO EN LOS NOVENTA

Los esfuerzos mundiales por llevar a cabo reformas educativas en las últimas dos décadas han estado motivados, principalmente, por un creciente interés político en asegurar la cobertura y el acceso universal a la educación, así como el aprendizaje efectivo de todos los alumnos de los nuevos contenidos curriculares. Este renovado interés por la calidad de la educación se inscribe en un contexto de fuertes cambios políticos y económicos que son interpretados en formas variadas por las diferentes corrientes académicas y de opinión.

Algunas voces destacan la existencia de condiciones en las que las reglas de libre mercado, la globalización y las nuevas tecnologías de información imponen una redefinición de las metas y objetivos de la educación pública, así como del rol del Estado para la consecución de esas metas. Estos cambios se caracterizan por una serie de objetivos que Carter y O'Neill (en Ball, 1998)

¹ Cabe añadir que la selección de los países estuvo condicionada por las limitaciones financieras de los investigadores, que realizaron las entrevistas personales y telefónicas haciendo uso de tiempo y recursos financieros de otros proyectos de investigación.

definen como constitutivos de una “nueva ortodoxia”. Algunos de esos objetivos serían: mejorar las economías nacionales estableciendo vínculos más robustos entre escolarización, empleo, productividad y mercado; mejorar los resultados de aprendizajes relacionados a las competencias y destrezas que requiere el mundo laboral; lograr un mayor y más directo control de los sistemas educativos nacionales sobre los contenidos curriculares y sobre la evaluación; reducir los costos de gobierno en educación; y ampliar la contribución de la comunidad a la educación, mediante su participación en la toma de decisiones escolares y la presión de la libre elección de mercado.

Estas tendencias vienen de la mano de nuevas formas de administración y gestión que plantean la necesidad de mejorar la eficiencia de los sistemas educativos y de fomentar la responsabilización social y profesional por los resultados de la educación. Estas formas de gestión, usualmente denominadas de “gerenciamiento”, suponen el uso de estilos y herramientas diferentes a los tradicionalmente utilizados por el sector público. Ejemplo de ello son la delegación de responsabilidad y poder de decisión a las instituciones proveedoras de servicios (escuelas), así como el control “a distancia” de la calidad en la provisión de esos servicios por parte del poder central. De esta manera, el Estado reemplaza la intervención directa y la prescripción por la determinación de objetivos, la comparación y la responsabilización (Lyotard, en Ball, 1998). Dentro de este marco de interpretación, sin duda, ha tenido un rol fundamental el uso de mediciones sobre el rendimiento académico de los estudiantes, que permite establecer mecanismos de responsabilización para los alumnos mismos, para los docentes y para las escuelas, en forma de graduación y promoción, acreditación, incentivo a la matrícula, mejoras salariales, y otras tantas formas de estímulo para elevar los resultados educativos.

Paralelamente, muchos países, entre ellos los latinoamericanos, buscan, por medio de las reformas curriculares y de gestión, fomentar y consolidar valores democráticos, de participación ciudadana y de respeto por la diversidad. En este sentido son numerosos los esfuerzos de muchos países de la región por establecer sistemas nacionales de medición de la calidad educativa que permitan, no necesariamente impulsar mecanismos punitivos o de mercado (*school*

choice), sino hacer más transparente y participativa la gestión mediante la provisión de información a los actores y usuarios del sistema. Visto de este modo, el Estado no se desentien- de de la gestión ni de su responsabilidad por promover mayor equidad y calidad de educación, sino que busca alentar a la población a que utilice la información sobre los resultados educativos para demandar localmente mayores niveles de calidad para sus escuelas.

Simultáneamente, el Estado se reserva para sí la función redistributiva de los ingresos públicos, que se puede hacer más efectiva y equitativa mediante evaluaciones que permitan distinguir cuáles son las subpoblaciones del sistema con mayores dificultades y menores recursos, y así poder intervenir con recursos materiales y pedagógicos para compensar las diferencias de logro académico. Adicionalmente, las evaluaciones han permitido reunir sistemáticamente mayores evidencias sobre las actitudes de los alumnos con relación a los aprendizajes mismos y a los valores personales y sociales, que son contenidos explícitamente delineados en la gran mayoría de los currículos de América Latina donde las reformas educativas de los últimos veinte años se han emprendido en contextos políticos de democratización y pacificación (Ferrer, 1999)².

Es importante notar que la búsqueda del control más directo sobre los contenidos curriculares se ha fortalecido en los países que han tenido sistemas educativos tradicionalmente descentralizados, mientras que en aquellos donde la prescripción curricular ha sido históricamente centralizada, los esfuerzos se han concentrado más bien en la flexibilización del currículo en el nivel local y escolar. En cualquiera de los dos escenarios, sin embargo, se percibe la necesidad de realizar evaluaciones centralizadas que permitan medir el logro académico de todos los alumnos bajo un mismo parámetro y con relación a contenidos curriculares a los que todos los estudiantes deberían poder acceder.

En este contexto, una de las herramientas principales de gestión educativa en muchos países del

² También en algunos países de Asia, como Japón y Corea del Sur, se ha hecho evidente la necesidad de complementar el enfoque tradicionalmente academicista de la educación con aspectos del desarrollo moral, afectivo y social de los estudiantes (GRADE, 2001).

mundo es la evaluación comparativa y cuantitativa de los logros de aprendizaje de los alumnos y la medición del impacto de diversas variables escolares y extraescolares sobre esos logros. Los países que disponían de sistemas nacionales de evaluación desde antes de los años ochenta, reforzaron su aplicación y los vincularon más directamente a sus nuevas metas curriculares. Los países que no disponían de tales sistemas de evaluación, como la mayor parte de los países latinoamericanos, los desarrollaron junto con el diseño e implementación de las reformas educativas de los años ochenta y noventa.

Opiniones especializadas argumentan que el impulso por la creación de estos sistemas de evaluación en los países en desarrollo estuvo influido por la presión de algunas agencias de cooperación internacional que condicionaron los préstamos para educación al desarrollo de sistemas nacionales de medición que permitieran medir los resultados de esas millonarias inversiones en reformas educativas. Sin embargo, esta interpretación es insistentemente desafiada por quienes consideran que, independientemente de las políticas de las últimas dos décadas, comúnmente denominadas “neoliberales”, la evaluación estandarizada del rendimiento académico y el estudio de los factores escolares y extraescolares que afectan ese logro fortalecen los cimientos democráticos de los sistemas educativos, en tanto permiten la opinión informada y la demanda de los beneficiarios, a la vez que el establecimiento de mecanismos de compensación para asistir a las comunidades escolares que más lo necesitan por parte del Estado.

A esta discusión se suma el hecho de que los cambios económicos mundiales de las últimas dos décadas también han dado lugar a una creciente preocupación de los países por su capacidad para competir eficazmente en los mercados internacionales. Para poder competir y prosperar, según se argumenta desde la publicación de *La naturaleza y las causas de la riqueza de las naciones* (Smith, 1776), es menester invertir en capital humano. Desde entonces, la educación ha sido vista como el principal medio para incrementar ese capital y fomentar así el desarrollo y la posibilidad de competir comercialmente en el escenario global.

La responsabilidad que se le adjudica a la educación por incrementar el capital humano de las naciones se manifiesta actualmente en las

múltiples iniciativas de reforma educativa en el mundo. Casi todas estas reformas, si bien muestran grandes diferencias, guardan como principal semejanza los esfuerzos por actualizar permanentemente los currículos y por medir, como se explicó anteriormente, la implementación y aprendizaje efectivo de los contenidos de esos currículos.

Comparaciones internacionales

Junto con el desarrollo y perfeccionamiento de sistemas nacionales de medición de aprendizajes, ha crecido notablemente el interés por las comparaciones internacionales de logro académico. Según Bempechat *et al.* (2000) “la fascinación con el logro académico en diferentes naciones ha sido motivada por los rápidos avances tecnológicos que han transformado la economía mundial [...] Los Estados-nación expresan su preocupación por su capacidad para competir en un mundo que se vuelve más y más pequeño”.

El caso de Estados Unidos es paradigmático en cuanto al impacto que las pruebas internacionales comparadas pueden tener sobre las políticas educativas de una nación. Los resultados en Matemáticas y Ciencias, de los estudiantes norteamericanos en pruebas internacionales, por debajo del promedio mundial y muy lejos de las potencias asiáticas, alarmaron a los líderes políticos de los inicios de los ochenta, quienes no dudaron en calificar a los Estados Unidos como “una nación en riesgo” (National Commission on Excellence in Education, 1983). La declaración hacía referencia al peligro que corrían los Estados Unidos de perder liderazgo mundial si su sistema educativo público no lograba mejores resultados de aprendizaje en Ciencias y Matemáticas, y desencadenó una serie de medidas de reforma curricular y de gestión en el país, con el objetivo explícito de ocupar el primer puesto en los *rankings* internacionales de logro académico hacia el año 2000.

Pero a las comparaciones internacionales no solamente se les atribuye un valor competitivo, o de *ranking*. Existen también argumentos conceptuales que justifican las comparaciones entre países como un instrumento metodológico que ofrece mayor información que la que puede ofrecer un análisis intranacional. Según Joseph Farrell (1979) no se puede realizar ninguna generalización científica sobre educación si no es a través de un estudio comparativo. El autor sostiene que

“los datos comparativos son esenciales para establecer la credibilidad de las proposiciones” y que “hasta que no se conoce la relación entre universos particulares y generales, ningún estudio es suficiente para probar una proposición”.

Para muchos investigadores y tomadores de decisiones sobre políticas educativas la participación en pruebas internacionales constituye una “intervención en sí misma”, independientemente de los resultados del estudio, y alerta a educadores e investigadores sobre problemas que podrían no haber considerado anteriormente.

Otras visiones sostienen que las pruebas internacionales pueden cumplir un rol más activo o propositivo. Husén (1987), cuando discute el impacto de la IEA (International Association for the Evaluation of Educational Achievement) sobre las políticas educativas en los países participantes, distingue dos categorías de “clientes”: a) formuladores de políticas (principalmente en el nivel central), y b) ejecutores de esas políticas (técnicos y educadores en el nivel central y local). Según Husén, los primeros suelen estar más interesados en datos y hallazgos de investigación referidos a la estructura del sistema de educación, mientras que los últimos prestan más atención a la información recogida sobre procesos de enseñanza-aprendizaje, prácticas y metodología de aula, materiales educativos y evaluación de aprendizajes en el aula.

Otros autores proponen algunos objetivos específicos que suelen perseguirse mediante la aplicación de pruebas internacionales. Theisen *et al.* (1983), por ejemplo, citan tres objetivos principales para esa participación: comparar el logro académico relativo entre países y por asignatura; identificar implicancias de política educativa para un país a través de los determinantes de logro hallados en otros países; y reevaluar las prioridades de inversión dentro de un país para elevar la calidad de los aprendizajes.

La teoría sobre educación comparada es extensa y controvertida, y los argumentos sobre el valor y validez de las comparaciones internacionales aún son materia de acaloradas discusiones en los medios académicos. Sin embargo, persiste el interés de los investigadores sociales, así como de los formuladores de políticas públicas alrededor del mundo, en realizar comparaciones, cada vez más precisas y sofisticadas, sobre diferentes aspectos de la política so-

cial. Este ha sido el caso de las pruebas internacionales de logro académico, que se siguen aplicando en una creciente cantidad de países cada año y que se van perfeccionando gradualmente en lo que se refiere a la especificación consensuada de sus objetivos, así como a las técnicas de recojo y análisis de la información.

I.2. ¿EN QUÉ CONSISTEN LAS PRUEBAS INTERNACIONALES?

Las pruebas internacionales comparadas son instrumentos de medición estandarizados que se aplican en varios países simultáneamente, en grados o grupos etarios seleccionados, y que recogen información sobre logros de aprendizaje conceptual y procedimental en diferentes disciplinas, y sobre los factores contextuales que se presume influyen en el logro académico de los estudiantes.

Las pruebas suelen incluir ítemes de opción múltiple, preguntas abiertas y/o demostraciones prácticas, como por ejemplo de laboratorio en el caso de las Ciencias Naturales. Entre los factores contextuales más importantes que se suele estudiar están la inversión per cápita en cada sistema nacional; las características familiares y socioculturales de los alumnos; la infraestructura escolar; la preparación académica de los docentes; la cobertura curricular efectuada en clase en las áreas disciplinares evaluadas; y las actitudes y valores de los alumnos con relación a esas áreas.

Las pruebas pueden ser curriculares o “acurriculares”, dependiendo de su grado de vinculación a los currículos de los países participantes. Cuando las pruebas son curriculares, los instrumentos se diseñan sobre una matriz común que representa los contenidos básicos que se ofrecen en los niveles educativos a ser evaluados en casi todos los países participantes. A estos contenidos comunes se suman contenidos considerados necesarios para una educación de calidad en el siglo XXI, según la opinión de profesionales y académicos internacionales especializados en las áreas disciplinares evaluadas. En el caso de las pruebas “acurriculares”, los diseñadores del estudio establecen estándares de contenido académico disciplinar que se consideran necesarios para una efectiva inserción social y laboral en el mundo de hoy, independientemente de los contenidos de los currículos de los países participantes.

Algunas de estas pruebas se han aplicado una única vez, a modo de diagnóstico o como ejercicio de investigación. Actualmente, se observa una tendencia al diseño de pruebas de aplicación cíclica para seguir el rendimiento de un grupo etario o grado escolar de estudiantes cada cierta cantidad de años, y así establecer relaciones longitudinales con respecto al desempeño académico de esas poblaciones.

Agencias de evaluación

Estas son pruebas diseñadas y coordinadas por agencias internacionales de evaluación, y con cierto grado de participación de las unidades de medición de la calidad educativa de los países. Algunas de estas agencias son privadas e independientes, de un perfil básicamente académico como es el caso de Educational Testing Services (ETS, 'Servicios de Evaluación Educacional') en Estados Unidos, o de la International Association for the Evaluation of Educational Achievement (IEA, 'Asociación Internacional para la Evaluación del Logro Académico'), que comenzó con un perfil similar pero que gradualmente fue convirtiéndose en una agencia de fuertes vinculaciones con los gobiernos nacionales participantes.

Otras agencias son intergubernamentales desde su concepción y en ellas están representadas las máximas unidades políticas o técnicas de administración o evaluación educativa de los países miembros. Ejemplo de estas últimas son la Organización para la Cooperación y el Desarrollo Económico (OCDE), el Laboratorio Latinoamericano para la Evaluación de la Educación (LLECE) –dependiente de Unesco-OREALC–, o el Southern Africa Consortium for Monitoring Educational Quality (SACMEQ, 'Consortio de África del Sur para la Evaluación de la Calidad Educacional').

Las pruebas de la IEA o de la OCDE son de cobertura mundial, mientras que otras, como las del LLECE o SACMEQ, son de cobertura regional. En la actualidad existen no menos de diez pruebas internacionales que miden el logro estudiantil en diferentes áreas de conocimiento, tal como se resume en el Cuadro 1.

Difusión y uso de los resultados

Independiente de su origen, estrictamente académico o de aplicación política, las pruebas in-

ternacionales buscan el desarrollo de nuevas metodologías de investigación comparada y enfoques analíticos (Forster, 2000); actualmente, sin embargo, se tiende a desarrollar estas pruebas como un instrumento para conseguir insumos de información para el diseño y mejoramiento de políticas educativas orientadas a elevar la calidad de los aprendizajes. Para ese fin, algunas pruebas ofrecen lo que en este informe llamaremos "opciones nacionales", es decir, la posibilidad de medir subpoblaciones específicas que por alguna razón requieran una medición más enfocada, o de áreas de conocimiento particulares y diferentes de las que evalúa la prueba general. Así, un país podría hacer uso del operativo internacional para evaluar, por ejemplo, un grupo etario diferente o agregar ítems de contenidos curriculares nacionales no contemplados en la prueba internacional.

Los resultados de las pruebas internacionales pueden ser reportados y difundidos por las agencias que dirigen los estudios, por las agencias nacionales que coordinan los estudios en los países participantes, por investigadores independientes o por la prensa (Forster, 2000; Macnab, 2000). Los datos que las pruebas arrojan suelen ser cuantiosos y por esa razón se eligen diferentes formas de reportarlos, de acuerdo a las audiencias a quienes están dirigidos. Algunos informes presentan un *ranking* de acuerdo al logro relativo de los países participantes, con relación al total del examen o a grupos de ítems seleccionados para el informe; también suele mostrarse la distribución de los resultados en relación a la media aritmética de cada país (dispersión), el logro de diferentes subpoblaciones, o las relaciones entre logro académico y diferentes variables contextuales relevadas por el estudio (Forster, 2000).

Algunas de las formas en las que se suele usar la información recogida son: la motivación para mejorar políticas a partir de la simple comparación con el logro de otros países; la promoción de políticas de mejoramiento a través de la difusión masiva de los resultados; el establecimiento de nuevos estándares académicos de acuerdo al desempeño de alumnos en otros países; la identificación de los factores escolares y extraescolares que muestran un efecto positivo en el logro de aprendizajes y el estímulo para las investigaciones dentro del país mediante la profundización de los resultados del estudio internacional (Forster, 2000; Macnab, 2000).

CUADRO 1

RESUMEN DE LAS PRINCIPALES PRUEBAS INTERNACIONALES APLICADAS EN AMÉRICA LATINA

Agencia	Prueba	Años de aplicación	Cobertura	Población objetivo	Áreas curriculares y competencias	Tipo de prueba	Principales factores asociados	Nº de países latinoamericanos 1/
OECD; Statistics Canada; NCES; ETS	IALS (International Adult Literacy Survey)	1994-1998	Global	16-65 años.	Lectura; prosa, documentación e información cuantitativa.	Preguntas cerradas y abiertas.	Variables culturales, sociales y demográficas	1/22
OECD; Statistics Canada; NCES	ALL (Adult Literacy and Lifeskills)	2002	Global	16 a 65 años.	Lectura; alfabetización numérica; razonamiento lógico; trabajo en equipo.	Preguntas cerradas y abiertas.	Variables sociales, educacionales, y económicas. Acceso y uso de NTC.	3/22 [Argentina, Bolivia y Costa Rica]
OECD	PISA (Programme for International Student	2000 2001	Global	15 años	"Destrezas para la vida": Matemáticas, lectura y Ciencias.	Opción múltiple y respuesta abierta.	Variables demográficas, sociales, económicas y educacionales.	6/33 [Argentina, Brasil, Chile, México, Perú y Uruguay]
LLECE (Unesco OREALC)	Primer Estudio Internacional Comparativo sobre Lenguaje, Matemática y Factores Asociados	1997	A. Latina	Grados 3 y 4	Lenguaje y Matemática.	Opción múltiple y preguntas abiertas en Lenguaje.	Variables escolares, familiares y sociales.	12 [Argentina, Bolivia, Brasil, Chile, Colombia, Cuba, Honduras, México, y Venezuela]
IEA	PIRLS (Progress in International Reading Literacy Study)	(RLS-1991) 2001+	Global	Grado 4 (9 años)	Lectura	Opción múltiple y respuestas abiertas.	Hogar; escuela; prácticas pedagógicas; currículum; políticas.	2/35 [Argentina y Colombia]
IEA	TIMSS (Third International Mathematics and Science Study) TIMSS-R	1995 1999	Global	Grados 3-4, 7-8, y último de educación secundaria.	Ciencias y Matemáticas.	Opción Múltiple, Respuesta abierta y demostración práctica.	Curriculm; Prácticas pedagógicas; políticas; actitudes; hogar; escuela.	2/45 [Colombia y México] 2/38 [Argentina y Chile]
IEA	Educación Cívica	1999 y 1999-2000 (dos fases)	Global	14 y 17-18 años	Ciudadanía y democracia; identidad nacional y relaciones internacionales; cohesión social y diversidad; economía y medio ambiente	Preguntas cerradas y abiertas; escalas de actitud en el hogar.	Clima escolar, actividades extraescolares, recursos culturales	2/32 [Colombia y Chile]

1/ Se indica el número de países latinoamericanos participantes sobre el número total de países participantes.

I.3. EL DEBATE ACERCA DE LA PARTICIPACIÓN EN PRUEBAS INTERNACIONALES

El valor de las comparaciones de logro académico entre países o entre regiones es materia de discusión permanente entre investigadores. Los argumentos que se esgrimen a favor y en contra de esas comparaciones suelen ser de carácter eminentemente técnico, aunque también entran en juego factores de orden político e ideológico. A continuación se resumen los beneficios y problemas más frecuentemente citados por la literatura especializada con relación a la participación en pruebas internacionales.

Beneficios

- Los resultados comparados internacionalmente atraen la atención de los medios de comunicación, movilizan fuerzas políticas (Greany y Kellaghan, 1996) y suscitan debates sobre las implicancias prácticas de los resultados obtenidos (Husén, 1987; Husén y Tuijnman, 1994, en Kellaghan, 1996).
- La construcción de pruebas requiere de los países participantes una revisión cuidadosa de sus currículos y su comparación con los de otros países, lo cual puede conducir a su actualización y mejoramiento (Beaton *et al.*, 1999). También motiva a educadores e investigadores a estudiar problemas del sector que no se había considerado antes (Husén, 1987).
- Se desarrolla la capacidad local para la medición de aprendizajes (Greany y Kellaghan, 1996; Torney-Purta, 1990; Bloom, 1974), en la medida en que educadores e investigadores de la educación se ven llamados a interactuar con especialistas internacionales y así se mantienen actualizados (Beaton *et al.*, 1999; Mislevy, 1995).
- Se conoce la efectividad de las políticas implementadas por otros países (Beaton *et al.*, 1999).
- Además del logro académico, permiten comparar otra gran variedad de variables educativas, tales como el clima escolar o los factores familiares y sociales que afectan el rendimiento académico (Husén, 1987; Torney-Purta, 1990). Si bien algunos de estos factores pueden caer fuera de las posi-

bilidades sectoriales de acción, todos contribuyen a una comprensión más integral del estado actual del sistema educativo y sus condicionantes.

- Los gobiernos pueden incluir en estos estudios preguntas e ítemes que de otra manera deberían evaluar por su propia cuenta, a través de sus sistemas nacionales (Husén, 1987).
- La dispersión de los estudios comparativos internacionales suele ser mucho más fuerte y evidente que la de los estudios nacionales, por lo que permiten indagar más profundamente en los efectos de diferentes variables sobre los resultados de aprendizaje (Husén, 1987).
- La construcción y la aplicación colectiva de pruebas puede promover prácticas profesionales democráticas debido al tipo de procedimientos y metodologías utilizadas (Torney-Purta, 1990).
- Algunas pruebas, como TIMSS, utilizan diseños jerárquicos para la recolección de datos, lo cual permitiría, asumiendo que exista suficiente varianza en los datos totales, analizar cuánto de la varianza se debe a diferencias entre países o a diferencias entre estudiantes dentro de cada país (Bempechat *et al.*, 2000). Estos análisis pueden contribuir a revelar cómo algunas políticas similares pueden tener efectos diferenciales en los países, dependiendo de otras variables complementarias³.
- Si bien en el proceso participan técnicos internacionales, los informes de resultados son producidos por investigadores nacionales que conocen a fondo los problemas de su país; por esta razón se espera que exista una interacción entre investigadores y educadores por lo menos igual a la que puede existir en el caso de pruebas nacionales, para que la información recabada y difundida provoque respuestas constructivas a los problemas detectados por el estudio (Bloom, 1974).

³ Un ejemplo de ello es el efecto positivo que la tarea en el hogar puede tener en el logro académico de los estudiantes chinos, mientras que en Perú el efecto puede ser mucho menor debido al alto índice de analfabetismo de los padres.

Problemas

- Los resultados de los estudios, en particular cuando se presentan como *ranking* de logro entre países, suelen desatar comentarios y críticas por parte de los políticos, los medios de comunicación y el público, en los que solo importa la posición relativa de logro en el escenario internacional y no la abundante información sobre factores sociales, escolares y personales asociados al rendimiento (Greaney y Kellaghan, 1996). La prensa, particularmente, tiende a desconocer los aspectos más complejos y auténticamente informativos de los resultados y solo destaca las posiciones en el *ranking*, a menudo con títulos dramáticos y escandalosos (Husén, 1987; Kellaghan, 1996).
- Si bien las pruebas internacionales utilizan diferentes formatos, algunos de ellos innovadores, los ítemes más numerosos continúan siendo diseñados en forma de opción múltiple. Aun cuando estos ítemes, si están debidamente contruidos, pueden evaluar habilidades cognitivas complejas, persisten críticas y dudas sobre su validez para ese fin (Greaney y Kellaghan, 1996).
- Se ha criticado a los estudios internacionales la falta de metodologías que fomenten la construcción cooperativa y democrática de las pruebas, lo cual pone en riesgo la participación efectiva y el beneficio que puedan derivar de esa participación los países con escaso desarrollo técnico (Husén, 1987)⁴.
- La complejidad de los estudios, sumada a la escasa capacidad que suele tener el público para interpretar los numerosos datos que derivan de los estudios comparativos, determina que la información que finalmente se difunde resulte ambigua o contradictoria, que un mismo dato pueda ser utilizado para argumentos antagónicos (dependiendo de la posición política), o que sencillamente se malinterpreten los datos y se llegue a conclusiones erróneas (Husén, 1987). Los indicadores, interpretados aisladamente de su contexto cultural y estructural, pueden conducir a interpretaciones y acciones de política equivocadas (Burstein *et al.*, 1992, en Schmidt y McKnight, 1995).
- La validez de los estudios se ve amenazada por las grandes diferencias culturales entre los alumnos de diferentes países, que hacen que las “condiciones” en que se toman las pruebas no sean iguales en cada país. Entre estas diferencias destacan los diversos grados de familiaridad que los alumnos puedan tener con las pruebas estandarizadas, los estilos de respuesta (como la tendencia a omitir o a adivinar respuestas), y las motivaciones que puedan tener los alumnos para hacer un esfuerzo significativo por resolver las preguntas (Mislevy, 1995).
- A menudo se toman decisiones de política sobre la base de interpretaciones de causalidad entre las variables estudiadas, cuando en realidad esa causalidad solo podría confirmarse mediante estudios experimentales controlados (Platt, 1975, en Mislevy, 1995).
- La información cualitativa recogida por los estudios comparativos no siempre resulta suficiente para establecer las variaciones que existen al interior de cada cultura y así comprender mejor cómo las creencias y percepciones de los alumnos afectan diferencialmente su rendimiento (Bempechat *et al.*, 2000).
- Si bien se ha propuesto el uso de las pruebas internacionales para fines de responsabilización de los gobiernos y funcionarios, aún no existen definiciones claras sobre quién será responsable, o quién tendrá el poder o la autoridad para modificar y mejorar el rendimiento académico de los alumnos. En general, y como sucede con muchos sistemas de evaluación nacionales, se ha dado poco espacio a la discusión sobre el uso exacto que se dará a las pruebas. Puede pretenderse, erróneamente, que la misma prueba sirva múltiples propósitos cuando en realidad esos propósitos pueden ser incompatibles (Kellaghan, 1996).
- Hasta la fecha, ni las agencias evaluadoras ni académicos especializados han llevado a cabo un registro sistemático de los usos que

⁴ Debe considerarse que esta construcción cooperativa tiene que ver tanto con la participación efectiva de los técnicos de todos los países involucrados, como con la posible participación de diferentes grupos profesionales en los ámbitos educativos nacionales para la validación de los contenidos de las pruebas.

se ha dado en diferentes países a los resultados de las evaluaciones. Kellaghan (1996) resalta que los informes que existen actualmente presentan información limitada y relativamente superficial, y por lo tanto no se pueden identificar los mecanismos exactos por medio de los cuales la información [de la IEA] se “traduce” en acciones de política. Aunque este no sea un problema inherente a las pruebas mismas, históricamente demuestra que las evaluaciones internacionales no siempre logran informar eficazmente a las audiencias o clientes principales.

El balance entre las ventajas y limitaciones que presentan las pruebas internacionales pareciera indicar que la participación en estos proyectos puede ser valiosa para los países, inclusive para los países en desarrollo (Greaney y Kellaghan, 1996), siempre y cuando se tomen ciertas previsiones con respecto a las características técnicas de las pruebas y a las condiciones políticas e institucionales para su aplicación, incluyendo mecanismos para la correcta interpretación y difusión de los resultados. Kellaghan (1996) propone una variedad de “condiciones” bajo las cuales debiera concretarse la participación de un país en pruebas internacionales comparadas. Esas condiciones son tomadas en cuenta para sustentar algunas de las recomendaciones que se efectúan hacia el final de este documento.

II. MARCO CONCEPTUAL Y CASOS DE ESTUDIO

El presente estudio fue estructurado sobre un marco conceptual que encuadra la participación en pruebas internacionales de aprendizaje en una secuencia de cuatro fases:

Iniciativa y adopción. Los resultados de aprendizaje obtenidos por un país, según muestran las mediciones nacionales de rendimiento u otros estudios académicos de investigación educativa, estimulan la opinión y la investigación sobre la necesidad de elevar la calidad de los aprendizajes y sobre las formas de lograrlo. La unidad central de administración educativa, o alguna entidad académica con el aval oficial, resuelve que es necesario medirse con respecto al logro académico de otros países y, de acuerdo a las prioridades curriculares y de política educativa en cada país, se opta por uno o más tipos de proyecto de evaluación internacional.

Se espera de esta manera realizar un seguimiento de los alcances y efectos de los programas curriculares nacionales.

Implementación. En cada país se ponen en marcha los procesos técnicos de evaluación internacional, desde las primeras reuniones de discusión entre expertos sobre el marco conceptual a ser utilizado, hasta la selección y validación de ítemes, la selección de la muestra poblacional y la aplicación final de los instrumentos. Si la relación entre los equipos técnicos de medición y los profesionales del desarrollo curricular en cada país es fluida y cooperativa, es probable que el proceso de diseño de las pruebas contribuya a las primeras revisiones críticas de los contenidos curriculares nacionales, ya sean antiguos o de reciente diseño e implementación. Entran en juego en esta fase los aspectos financieros, así como las capacidades organizacionales y profesionales que afectan la viabilidad y calidad de los operativos de diseño y aplicación de las pruebas. En el contexto latinoamericano juegan un rol particularmente importante los aprendizajes técnicos, de tipo estadístico y psicométrico, que se desarrollan durante la experiencia.

Interpretación y diseminación de resultados. Se lleva a cabo la lectura de los resultados de evaluación por parte de los técnicos de medición así como de los profesionales del desarrollo curricular convocados durante el proceso de implementación. Esa lectura supone el análisis comprensivo y la valoración inicial del nivel de logro obtenido. Se preparan informes de resultados para la prensa, para la opinión pública y para diferentes audiencias profesionales del sector; se ofrecen talleres de difusión y desarrollo de capacidades de interpretación y análisis de los datos, y se liberan las bases de datos para uso de los medios académicos con fines de investigación. Los esfuerzos de interpretación se centran en las inferencias de tipo curricular que se pueden realizar con relación a los contenidos prescritos y enseñados de las áreas de conocimiento evaluadas.

Decisiones, impacto político y nuevas líneas de investigación. Los resultados de las evaluaciones influyen directamente o indirectamente mediante la presión de la opinión pública y profesional, sobre las decisiones de funcionarios y profesionales del sector para elevar la calidad de los aprendizajes.

Casos de estudio

Del total de pruebas internacionales de aprendizaje en las que América Latina ha participado o está participando actualmente, se decidió en este estudio reportar las siguientes experiencias:

- **Primer Estudio Internacional Comparativo sobre Lenguaje, Matemáticas y Factores Asociados del LLECE** (Laboratorio Latinoamericano para la Evaluación de la Calidad de la Educación), realizado bajo la dirección de la Oficina Regional para América Latina y el Caribe (OREALC) de la Unesco. Resulta interesante por tratarse del único estudio comparado de cobertura estrictamente regional, el cual contó con la participación de un número importante de países latinoamericanos.
- **TIMSS (Third International Mathematics and Science Study); TIMSS-R (repetición de TIMSS); y Educación Cívica.** Todas estas pruebas son administradas por la International Association for the Evaluation of Educational Achievement (IEA). Fueron incluidas por tratarse de pruebas que han convocado la participación de algunos países de la región y porque la experiencia de su aplicación puede ser contrastada con la de Estados Unidos, donde el impacto de esta evaluación ha sido mucho más significativo que en América Latina.
- **Pruebas PISA (Programme for International Student Assessment)** de la Organización para la Cooperación y el Desarrollo Económico (OCDE). Fue elegida debido a que es una prueba “acurricular” que propone estándares internacionales y ha concitado el interés de varios países de la región.

III. RESULTADOS DEL TRABAJO DE CAMPO

III.1. AGENCIA: LLECE (LABORATORIO LATINOAMERICANO PARA LA EVALUACIÓN DE LA CALIDAD EDUCATIVA, DE UNESCO-OREALC) - PRIMER ESTUDIO INTERNACIONAL COMPARATIVO SOBRE LENGUAJE, MATEMÁTICAS Y FACTORES ASOCIADOS

Cuando se aprobó la idea de crear el Laboratorio Latinoamericano para la Evaluación de la Calidad Educativa - LLECE (en adelante, el Laboratorio), prácticamente todos los países de la región optaron por adherirse. Fue concebido como una red

de los sistemas nacionales de evaluación de principios de los noventa, en un momento en que la responsabilidad de crear y sostener un sistema de evaluación era complejo desde el punto de vista técnico y político. El Laboratorio ofrecía un espacio de diálogo y participación para los funcionarios y técnicos responsables de esos sistemas y brindó apoyo para la creación de varios sistemas nacionales de evaluación.

Las principales motivaciones para realizar el estudio comparado fueron la necesidad de establecer una comparación “de raíz latinoamericana” que contemplara las particularidades culturales y educativas de la región y obtener información para apoyar las decisiones de política educativa. Otra razón que motivó a los países a escoger la prueba del Laboratorio fue que era mucho más económica que otros estudios internacionales.

El Primer Estudio Internacional Comparativo sobre Lenguaje, Matemáticas y Factores Asociados fue aplicado en 1997 en 13 países latinoamericanos, en los grados 3º y 4º de educación primaria, en las áreas de Lenguaje y Matemáticas. La prueba incluyó ítemes de opción múltiple y también de respuesta abierta, y recogió información sobre variables asociadas al logro académico, entre ellas de contexto socioeconómico, familiar, escolar y áulico. Los principales objetivos del Laboratorio fueron entregar una comparación de rendimiento académico entre países y analizar los factores asociados a ese rendimiento. Objetivos complementarios fueron: fortalecer los sistemas nacionales de medición y evaluación por la vía de la práctica a través del primer estudio comparado; establecer estándares educativos, y generar una línea de investigación a partir de los resultados.

Los resultados confirmaron muchos de los datos que algunos sistemas nacionales ya habían comenzado a producir, como por ejemplo las diferencias de logro a favor de las escuelas urbanas y privadas, escasas diferencias de rendimiento entre géneros, y una fuerte asociación positiva entre bajo nivel socioeconómico y bajo rendimiento académico. En cuanto a los resultados comparados internacionales, fue particularmente llamativa la diferencia entre Cuba y el resto de los países de la región, con más de dos desviaciones estándar a favor de Cuba. Los resultados fueron publicados por LLECE en dos informes. El primero presenta los resultados

comparativos en forma de *ranking* y desagregados por estratos. El segundo informe, que salió tiempo después, ofrece un análisis de factores asociados al rendimiento académico en el conjunto de países participantes, y se concibió con la idea de ofrecer mejores insumos de información para la toma de decisiones de política educativa y para estudios académicos en la región.

El estudio fue financiado con aportes del Banco Interamericano de Desarrollo (BID), sumados a algunos fondos de la Fundación Ford y dinero de los países participantes el cual se les devolvía en forma de servicios de capacitación para la medición.

Para garantizar que los ítemes de las pruebas tuvieran una fuerte vinculación con los currículos de los países participantes, se invirtieron dos años y medio en el diseño de las mismas, con reuniones para la selección de ítemes, verificando currículos nacionales de Matemáticas y Lenguaje. Este fue uno de los aspectos más beneficiosos del proceso, pues obligó a los países participantes a efectuar una importante revisión de sus currículos, lo que les permitió descubrir sus debilidades y fortalezas relativas.

Los representantes ante el Laboratorio, llamados “coordinadores nacionales”, fueron en su mayoría los directores de los sistemas de evaluación de cada país, o de la oficina ministerial de la cual esos sistemas dependían. En algunos casos, el diseño de la matriz curricular se hizo con la participación de los técnicos en currículo de los diferentes ministerios. Sin embargo, en algunos países hubo tensiones importantes entre las oficinas del currículo y los sistemas nacionales de evaluación, lo cual determinó que los departamentos de currículo no estuvieran representados durante el proceso. Debido ello, se reconoció que la prueba final no tendría una base curricular de igual significado para todos los países, y que ello repercutiría en las inferencias de orden curricular que se hicieran más adelante.

La entrega de los resultados finales por parte de la Unesco-OREALC se hizo directamente a los ministerios. Al mismo tiempo, se realizaron conferencias de prensa en Santiago de Chile y se editó un compacto para televisión que fue difundido en América Latina por la Red de TV Española. Aunque todos los países participantes de la prueba del Laboratorio se habían comprometido a publicar los resultados internacionalmente, dos de los países participantes, México y

Perú, cuestionaron la validez de los resultados y desautorizaron su difusión.

Era potestad de los gobiernos decidir su difusión en cada país. De ahí que, en muchos casos, los resultados solo circularan en reuniones de cúpula, acaso compartiendo un resumen con la prensa, pero sin la voluntad de generar debate público sobre el tema.

Los responsables por la ejecución del estudio en el Laboratorio coinciden con algunos de los coordinadores nacionales en que la falta de una estrategia fuerte de difusión determinó que los resultados de la prueba no tuvieran un gran impacto político, a través de la opinión pública, ni en las decisiones de política educativa de la región. Las reacciones por parte de los gobiernos fueron variadas, dependiendo de las expectativas y de la coyuntura política de cada país, aunque se reconoce que en general, exceptuando Cuba, fueron de “tono bajo”. Sin embargo, se esperaba que con la difusión del segundo informe sobre factores asociados y con el paso del tiempo, los datos comienzen a ser utilizados por investigadores nacionales e internacionales para análisis secundarios y proyectos de investigación, probablemente comisionados por agentes no gubernamentales, así como por académicos y organizaciones sindicales.

III.2. AGENCIA: IEA - TIMSS, TIMSS-R Y EDUCACIÓN CÍVICA

La prueba **TIMSS** administrada por la IEA fue aplicada en 1995 en unos 45 países de todo el mundo. En algunos países solo participaron algunas provincias y no todos evaluaron todas las poblaciones objetivo del estudio: los grados 3° ó 4° de primaria, 7° u 8° grados, y último año de la educación básica. Las pruebas se centraron en las áreas de Matemáticas y Ciencias y la matriz curricular se definió mediante un relevamiento previo de los contenidos curriculares prescritos tanto por documentos oficiales como por libros de texto y materiales didácticos de uso masivo en un número significativo de países. La prueba incluyó ítemes de opción múltiple, de respuesta abierta y de demostraciones prácticas (por ejemplo, de laboratorio). También se realizó un estudio cualitativo, mediante videograbaciones, sobre procesos pedagógicos de aula en seis países de la muestra.

En cuanto a los resultados comparados internacionales, quedó confirmado que los países del

CUADRO 2

RESUMEN DE LA EXPERIENCIA DE CUATRO PAÍSES EN LA PRUEBA DEL LLECE

	Aprendizaje técnico y críticas	Difusión de resultados	Uso de los resultados e impacto en la política educativa
Argentina	<p>Se valora la oportunidad de compartir experiencias con América Latina y de discutir con otros países en situación de desigualdad técnica –para lo cual se requiere mucha tolerancia y buena voluntad– y de llegar a acuerdos para la toma de decisiones.</p> <p>Crítica: se tiene la percepción que el Laboratorio no supo aprovechar adecuadamente los conocimientos y la experiencia que ya tenían algunos países</p>	<p>La principal receptora de los resultados fue la ministra de Educación, quien los compartió con el gabinete ejecutivo de la Presidencia. La lectura dada a los datos fue mínima, en parte porque el primer informe solo fue comparativo y no permitía realizar inferencias sobre las razones que explicaban esos resultados. No se previeron estrategias de difusión y a la prensa solo se le entregó una copia del informe oficial del Laboratorio, sin interpretaciones por parte del Ministerio.</p>	<p>Escasa repercusión a nivel del Ministerio de Educación, pues los resultados llegaron poco tiempo antes de una renuncia de todo el equipo de alta dirección. Sin embargo, muchos de estos ex funcionarios que tuvieron acceso a la información circulan actualmente en medios académicos, desde donde tienen cierta influencia en la formación de profesionales y especialistas en educación.</p>
Colombia	<p>Producto del proceso, se valora el fortalecimiento de la relación entre el Ministerio de Educación y el ICFES (Instituto Colombiano de Fomento a la Educación Superior) que actuó como correpresentante ante el LLECE.</p> <p>Críticas: Poca participación efectiva del país en el diseño de ítemes, falta de acceso al marco conceptual con el cual se construyeron los instrumentos, poca comunicación entre los países participantes.</p>	<p>El gobierno no tenía recursos para publicar los documentos, y solo se entregaron copias del informe de OREALC a gobernadores o secretarios de educación de departamentos. Se piensa que los resultados no están “contextualizados” a la realidad regional, debido a la ausencia de estándares regionales y a que el proyecto no ofreció un marco interpretativo de los resultados con relación a los objetivos y alcances de las reformas y políticas educativas implementadas en las dos últimas décadas.</p>	<p>Débil impacto sobre la educación. Se considera que las conclusiones a las que llega el Laboratorio no están planteadas de forma que pueda afectar la toma de decisiones de política.</p> <p>Solo el ICFES tomó la iniciativa de profundizar el análisis de los datos.</p>
Cuba	<p>La experiencia fue valorada como de amplio aprendizaje para diversos actores del área educativa. La participación en el diseño de las matrices curriculares permitió identificar falencias del currículo cubano. El debate y la confrontación técnica entre diferentes formas de evaluar permitió acceder a nuevos conocimientos sobre evaluación, como los relativos al diseño y calificación de respuestas abiertas y nuevas tecnologías de medición (modelo de Rasch y modelos jerárquicos lineales).</p>	<p>Se dio amplia difusión a los resultados del primer informe, con un marcado y deliberado cariz político, destacándolo como un logro significativo de la política social cubana.</p> <p>Se recibió con beneplácito la información sobre los efectos positivos del clima escolar sobre los aprendizajes, contenidos en el segundo informe.</p>	<p>Atendiendo a los buenos resultados, se decidió premiar a los docentes con un incremento del 30% en sus salarios, medida que también pretendía desalentar el doble empleo de los maestros, pues los resultados del estudio mostraban una relación negativa entre el multiempleo y altos rendimientos estudiantiles. En el terreno curricular, se decidió trabajar con mayor diversidad de textos, por lo que se incentivó la producción de nuevos materiales educativos.</p>
Perú	<p>Se consideró de gran utilidad un curso de capacitación de 15 días que el Laboratorio ofreció en Santiago de Chile.</p> <p>Críticas: limitadas posibilidades de influir sobre el diseño de la prueba, no se percibe que la experiencia haya constituido un espacio de identidad y comunicación regional.</p>	<p>Los representantes peruanos pusieron objeciones técnicas a los resultados y solicitaron postergar su publicación internacional hasta obtener mayores detalles sobre el análisis de los datos. Sin embargo, desde antes de la conclusión del estudio, se presumía que el gobierno no tenía intenciones de publicarlos, por razones políticas. La no publicación de los resultados fue interpretada en el país como una privación al derecho constitucional de acceso a la información pública. Los resultados fueron liberados durante el Gobierno de Transición, mediante una publicación de la Unidad de Medición de la Calidad Educativa del Ministerio.</p>	<p>El impacto más evidente de la participación en el estudio internacional del Laboratorio fue quizá la tardía publicación de los resultados, que es considerada un símbolo importante de la democratización del sector. Además, se observa que diferentes medios y actores sectoriales citan frecuentemente los bajos resultados del Perú con respecto a la región para sustentar argumentos sobre la prioridad de la política educativa.</p>

sudeste asiático, como Japón, Singapur y Corea del Sur, logran resultados muy superiores a la mayoría de los países industrializados del mundo. El único país latinoamericano que participó y permitió la publicación de los resultados, Colombia, obtuvo uno de los últimos lugares en el *ranking* internacional.

En 1998 se aplicó **TIMSS-R**, una repetición de la prueba, en 38 países, incluyendo tres latinoamericanos. TIMSS será aplicado nuevamente en el año 2003.

La prueba de **Educación Cívica**, también administrada por la IEA, se aplica regularmente en países de todo el mundo. Los contenidos de la prueba se centran en temas de ciudadanía, gobierno y democracia. También se recogen datos sobre factores sociales, familiares y actitudinales asociados al rendimiento académico.

III.3. AGENCIA: OCDE - PRUEBA PISA

PISA, la prueba comparativa de la OCDE, se aplica regularmente en países de todo el mundo a estudiantes de 15 años de edad, cualquiera sea su grado de escolarización. Los países miembros de la OCDE ingresan a la prueba automáticamente, mientras que los demás países son invitados a sumarse al proyecto mediante el pago de una cuota de membresía. Las pruebas de PISA se diseñan a partir de un conjunto de estándares de contenido y desempeño, elaborados por un grupo de especialistas en diferentes áreas de conocimiento, que se consideran imprescindibles para la adecuada inserción de los jóvenes en la sociedad y en el mundo laboral de hoy. PISA volverá a aplicarse en el año 2003 y luego en el 2006.

Al igual que en el caso de las pruebas de la IEA, la decisión de participar en la prueba PISA la tomó cada país individualmente y por lo tanto no es posible presentar testimonios de representantes regionales para esta experiencia de evaluación. En consecuencia, se han tomado en cuenta las voces de los representantes nacionales de evaluación de Argentina, Chile, México, Perú y Uruguay.

IV. RESUMEN CRÍTICO DE LOS RESULTADOS

Como se ha podido observar, existe ya un conjunto importante de experiencias de participación en pruebas internacionales en América La-

тина. Esas experiencias muestran diferentes grados de inserción e impacto efectivos en el escenario de las políticas y la práctica educativa de cada país, dependiendo de los contextos políticos e institucionales en que se inscriben. Aunque el tamaño de la muestra de países en este estudio, y el número limitado de actores entrevistados, limitan las posibilidades de hacer generalizaciones amplias a la región, la cantidad de información obtenida merece un resumen crítico que reúna los aspectos más importantes de los casos analizados, acuerdo a las fases de evaluación comparada internacional utilizadas como marco conceptual en este estudio.

En cuanto al momento previo a la iniciativa y la adopción de pruebas, en la que perciben los primeros “descontentos” con respecto a los resultados de aprendizaje obtenidos por los sistemas educativos nacionales, en América Latina esa percepción se basa, al menos parcialmente, en algunos referentes empíricos, tales como los resultados de pruebas nacionales de logro académico, pruebas exploratorias comparadas en la región, o resultados de pruebas internacionales anteriores. Parte del “descontento” sobre la eficacia de los sistemas educativos, que conduce a la búsqueda de información comparada sobre rendimiento académico, deviene del convencimiento, no del todo fundamentado, de que los grandes problemas económicos de la región tienen origen, en parte, en las deficiencias de aprendizaje de los estudiantes, quienes se gradúan sin estar preparados para las demandas de conocimiento del mundo actual.

Ese uso de referentes empíricos es decididamente más fuerte en países como Estados Unidos, donde se han empleado en forma mucho más extensiva los datos recogidos en pruebas nacionales y pruebas internacionales anteriores, y donde el temor a perder competitividad en el mercado internacional es –o al menos lo fue durante la última década– más marcado.

Las **iniciativas de participación**, así como la elección del proyecto de evaluación comparada más conveniente para cada país en América Latina, han estado en la práctica definidas por tres objetivos principales:

- a) La oportunidad de acceder a capacitación técnica en medición de aprendizajes y diseño de instrumentos. Existe, sobre la base de experiencias anteriores, una expectativa y percepción compartidas entre los países

CUADRO 3

EXPERIENCIA DE CUATRO PAÍSES EN LAS PRUEBAS DE LA EIA: TIMMS, TIMMS-R, EDUCACIÓN CÍVICA

	Aprendizaje técnico	Difusión de resultados	Uso de los resultados e impacto en la política educativa
Chile (TIMMS-R y Educación Cívica)	En el caso de la prueba de Educación Cívica, se desarrollaron capacidades relacionadas con las estrategias operativas en el trabajo de campo, con diseños de manuales, aplicación de instrumentos y análisis estadístico. Importante aprendizaje sobre el modelo psicométrico IRT, el cual pudo ser aplicado luego a los operativos de evaluación nacionales.	Los resultados de TIMSS-R fueron publicados solo en forma de ranking por un diario nacional, mientras que la Unidad de Currículum y Evaluación preparó un documento de difusión resaltando los aspectos más sustantivos de los datos recogidos. Para el estudio de Educación Cívica, que incluyó una opción nacional, los resultados incluyeron todos los datos comparativos internacionales de base. Los datos de la aplicación diferenciada (ítems de opción nacional) fueron difundidos y analizados posteriormente. Los resultados de esta prueba se presentaron en una capacitación para 40 mil profesores.	Es notoria la cantidad de artículos escritos por los académicos de uno de los centros de investigación sobre políticas públicas de Chile, que utilizan los datos para hacer críticas y sugerencias concretas sobre la educación pública en el país. Desde el Ministerio de Educación se está intentando que la información llegue a los profesores, a través de instancias de perfeccionamiento docente, para que haya impacto en el aula y en los contenidos.
Colombia (TIMMS y Educación Cívica)	Fue una excelente oportunidad de aprendizaje para los académicos participantes, especialmente en la etapa de investigación del marco contextual de la educación en Colombia y en el desarrollo de criterios fundamentados para la calificación objetiva de las preguntas abiertas. En el caso de la prueba de Educación Cívica, se valora la oportunidad de participación en el establecimiento del marco conceptual que guiaría la construcción de las pruebas.	Difusión de los resultados de Cívica, a través de talleres y seminarios para las escuelas participantes. Hubo importante interés del Ministerio de Educación en la divulgación de los resultados de TIMSS, que incluyó un seminario previo a la publicación de los resultados para funcionarios del Ministerio y del ICFES y publicaciones especiales para maestros de Matemáticas con los resultados comparativos con los demás países participantes.	Sobre la base de los resultados, el Ministerio de Educación hizo publicaciones que ofrecían recursos pedagógicos para que los profesores logaran mejores resultados de aprendizaje. Uno de los grandes impactos del TIMSS fue que alentó a los técnicos del Servicio Nacional de Pruebas del ICFES a rediseñar el examen que regula el ingreso a la educación superior, con un marco conceptual de referencia a criterios en vez de normas, y con análisis de respuesta al ítem.
Estados Unidos (TIMMS, TIMMS-R en el caso de algunos estados)	En el nivel local, el aprendizaje más valioso obtenido a través de la participación en TIMSS fue el relacionado al análisis e interpretación de datos. En el nivel nacional, se destacan los avances teóricos logrados mediante el desarrollo de nuevos marcos conceptuales y metodologías de investigación y análisis.	Se publicaron primero los resultados del análisis curricular y varios meses después el ranking internacional. Esto contribuyó a que, cuando se supo que Estados Unidos ocupaba un lugar modesto en ese ranking, las audiencias más especializadas ya tuvieran un conocimiento importante sobre los aspectos sustantivos del estudio. Se contrató una prestigiosa consultora de prensa de Nueva York para las relaciones públicas y las estrategias de difusión de TIMSS. La consultora proveyó entrenamiento a los coordinadores nacionales y asignó personal especializado a la "colocación" de noticias sobre los análisis de TIMSS en diferentes medios de comunicación del país. También en el nivel local se realizaron importantes esfuerzos para lograr una mejor interpretación y difusión de los resultados.	Los resultados de TIMSS pusieron en marcha una serie de medidas gubernamentales, no gubernamentales y profesionales para establecer estándares curriculares más altos, y para alinear los diferentes elementos de la gestión y la práctica educativa desde una perspectiva sistémica. En el nivel local o distrital, los esfuerzos se centraron en la alineación curricular local con los estándares estatales y en mejorar las condiciones pedagógicas para la entrega de mejores oportunidades de aprendizaje. Un efecto singular de TIMSS en el nivel local fue la creación de SMART, consorcio que ha decidido proveer extensas oportunidades de capacitación a docentes sobre currículo, métodos instruccionales y, especialmente, sobre cómo analizar datos de pruebas estandarizadas.
México	Se lograron aprendizajes técnicos valiosos vinculados a técnicas de muestreo y de organización operativa para las aplicaciones.	El país invocó razones técnicas para desautorizar la publicación internacional de los datos.	Se conservaron las bases de datos con los resultados dentro del ámbito de la SEP, pero no fueron analizadas posteriormente.

CUADRO 4

EXPERIENCIA Y/O EXPECTATIVAS DE CUATRO PAÍSES EN LAS PRUEBAS PISA

	Aprendizaje técnico	Difusión de resultados	Uso de los resultados e impacto en la política educativa
Chile (En proceso para participación en PISA 2003)	Se percibe que, en la medida en que se gana experiencia en medición internacional, y a partir de la unificación de los equipos de evaluación y currículum bajo una misma unidad administrativa, la relación entre los procesos de evaluación y desarrollo curricular se va fortaleciendo, especialmente a través de los procesos de revisión curricular sistemática que exigen las pruebas internacionales.	Aun cuando Chile no participó PISA 2000, el Ministerio de Educación difundió públicamente los resultados y anunció que Chile participaría en la siguiente aplicación del proyecto. Se espera difundir los resultados de PISA en forma más beneficiosa y efectiva que con las pruebas internacionales anteriores en que el país ha participado, porque se contará con más tiempo para la planificación de una estrategia adecuada y con mayor apoyo por parte de las organizaciones internacionales.	El gran desafío es lograr que los resultados de las pruebas no causen solo reacciones críticas contra la administración educativa nacional sino que, además, sean utilizados para implementar mejoras pedagógicas y curriculares. Sin embargo, dado que los nuevos marcos curriculares recién están terminando de implementarse en los años superiores de la educación media, se calcula que los resultados de PISA no promoverán cambios en el currículo en el corto plazo.
México (Participación en PISA Plus, PISA 2001 y en preparación para PISA 2003)	La experiencia de evaluación en el proyecto PISA ha permitido un mayor grado de participación técnica que en experiencias anteriores de evaluación internacional La sección de preguntas abiertas de la prueba ha demandado de las autoridades de la SEP una convocatoria más amplia a actores del medio educativo en el proyecto, específicamente para llevar a cabo la calificación de respuestas.	Los primeros destinatarios de los resultados serán los equipos técnicos y autoridades políticas de la próxima administración gubernamental nacional, lo cual aumentará las posibilidades de que los datos sean utilizados y que el proyecto evaluativo se continúe en el tiempo. Respecto a los resultados preliminares de PISA, que consistieron en una aplicación específica de PISA Plus para que México se comparase con otros países de la región antes de ingresar en la aplicación general de 2001, la Subsecretaría de Planeación y Coordinación Institucional ya ha difundido un documento interno. Los datos y las bases mismas también se han entregado a las administraciones estatales.	Ya se están realizando, en función de los datos preliminares obtenidos, algunas propuestas para orientar mejoras en el Programa Nacional de Lectura y en el Programa de Escuelas de Calidad. También se han realizado algunas propuestas para el desarrollo curricular de formación docente inicial, en el currículo de Matemáticas de Educación Básica, y se han facilitado las bases de datos para su análisis a la Academia Mexicana de Ciencias y al Centro de Investigación y Docencia Económica de México.
Perú (Al momento del estudio, en proceso para participación en PISA 2003)	Se considera que esta experiencia brinda a los técnicos la posibilidad de ensayar con aspectos técnicos y conceptuales diferentes a los de las pruebas nacionales.	Se considera de suma importancia lograr que se difunda información sustantiva, no solamente el ranking internacional. Para un óptimo aprovechamiento de la información, la coordinación del estudio en Perú pretende realizar un taller donde se explique a actores clave del sector educativo cómo interpretar los resultados de PISA.	Se sabe que las noticias de prensa serán escandalosas, pero se cree que en la medida en que el país siga participando y el público se acostumbre a que hay otras informaciones rescatables además del ranking, los medios de comunicación deberán buscar otro tipo de noticias para vender a la población. Un aspecto positivo de la prueba es que el país será propietario de los instrumentos y podrá aplicarlos nuevamente si quiere profundizar las investigaciones.
Uruguay (En proceso para participación en PISA 2003)	Se perciben tres importantes posibilidades de aprendizaje con el programa PISA: 1) generará un mayor conocimiento para el diseño de ítems de las áreas que actualmente no son evaluadas por las pruebas nacionales; 2) permitirá desarrollar capacidades específicas con relación al diseño, calificación y procesamiento de la información de pruebas de respuestas abiertas, y mejor conocimientos de la teoría de respuesta al ítem; y 3) se aprenderá más sobre la medición de variables escolares, curriculares e instruccionales.	Se planea difundir ampliamente los resultados de la prueba, así como lo que pudiera revelarse del instrumento mismo, tal como se ha hecho con las pruebas nacionales. De esta manera, se podría explicar a cada equipo de profesores, en cada establecimiento, qué es lo que las diferentes actividades de la prueba miden, para que ellos puedan aplicar la prueba, o ítems seleccionados de la misma, y así comparar el desempeño de sus alumnos con el de los estudiantes en otros países.	Los técnicos nacionales de medición reconocen que, a pesar del buen uso que se pretende dar a las pruebas, la medición contra estándares internacionales como los de PISA conlleva un riesgo político considerable. Por lo tanto, antes de ingresar en una de estas pruebas, consideraron necesario contar con el aval de las autoridades políticas del sector para asumir el costo de salir en los últimos lugares del ranking internacional.

de la región en cuanto a los beneficios de participar en pruebas lideradas por agencias de alta capacidad técnica y organizacional como la IEA o la OCDE. Se valora especialmente lo que se puede aprender de la organización operativa de los proyectos, de la calidad técnica de los instrumentos de medición, de los procedimientos para garantizar la calidad y validez de la información y de los mecanismos que esas instituciones implementan para brindar apoyo y orientación a los países participantes. En cuanto a la vinculación de los proyectos comparativos con los sistemas nacionales de medición, se espera desarrollar capacidades técnicas que fortalezcan a estos últimos, pero no se ha diseñado generalmente un “cruce” sistemático entre los datos obtenidos en ambos tipos de mediciones.

- b) La posibilidad de obtener información objetiva sobre el logro académico de los estudiantes con relación a estándares internacionales de contenido, que impacte la opinión pública y facilite una toma de decisiones mejor informada para el mejoramiento del sistema educativo.
- c) La expectativa, al menos por parte de algunos actores, de que la participación en una prueba internacional estimule la propuesta de estándares regionales contra los cuales medir el rendimiento de los países latinoamericanos. Si esto último no es posible, se espera al menos participar en pruebas internacionales de cobertura mundial que ofrezcan un marco conceptual sólido que permita interpretaciones válidas para el contexto latinoamericano. Cabe destacar que, si bien muchas de las decisiones de participación en pruebas internacionales en América Latina han estado inspiradas en la idea de establecer estándares académicos más claros y exigentes y compararse con ellos, lo que demuestra un interés por los aspectos curriculares de las reformas, la mayor parte de las decisiones se ha tomado sin la participación directa de los responsables de las unidades de currículo de los organismos de gestión educativa nacionales, ni de otros actores que puedan ejercer influencia sobre el desarrollo curricular en cada país. En este sentido, se podría afirmar que las iniciativas y decisiones de participación en pruebas internacionales están solo parcial-

mente alineadas con los objetivos nacionales de reforma y logro curricular.

Si bien existen algunas opiniones “supranacionales” en la región que entienden la evaluación internacional dentro de un contexto de política educativa basada en la noción de “gerenciamiento”, en parte motivada por el condicionamiento de los organismos internacionales, la mayor parte de las decisiones de participación no parece haberse tomado en función de ese modelo de gestión. Es decir, los beneficios que se esperan de la información recogida no están relacionados necesariamente a las nociones de descentralización y responsabilización por los resultados, sino más bien al deseo de proveer mayores herramientas de gestión para un adecuado desarrollo e implementación curricular. Ello, pese a lo señalado líneas arriba respecto a la débil vinculación de las iniciativas internacionales con las unidades responsables del currículo.

En cuanto a la fase de **implementación** de las pruebas internacionales, se aprecian grandes diferencias entre la experiencia de participación en la prueba del Laboratorio, por un lado, y en las pruebas de la IEA y de la OCDE, por el otro. Esas diferencias, están relacionadas principalmente con la capacidad técnica, organizacional, financiera y operativa de las agencias que administran las pruebas. La IEA y la OCDE aventajan al Laboratorio en todas esas características, a pesar de que la prueba regional reportó también varios beneficios a los países participantes.

En lo que se refiere a los aspectos estrictamente curriculares, la participación en pruebas de la IEA se ha dado en condiciones de desigualdad con respecto a los países desarrollados, pues América Latina ha ingresado tarde a los proyectos de evaluación, muchas veces con escasas capacidades técnicas para la representación, o con un débil aval financiero e institucional por parte de sus gobiernos nacionales. Por estas razones, las posibilidades de que los países de la región influyan en las decisiones con respecto a los marcos conceptuales y curriculares han sido extremadamente reducidas.

En el caso de la prueba del Laboratorio, si bien la matriz curricular fue discutida con los países, la falta de un mecanismo de comunicación multidireccional no permitió lograr un consenso con respecto a un conjunto de estándares deseables ni a un marco conceptual común que permitiera

una interpretación más sustantiva y contextualizada de los resultados y, por lo tanto, la pertinencia curricular del proyecto se vio fuertemente debilitada. Recordemos que en las pruebas de la IEA o de la OCDE en América Latina se valora precisamente que los países se midan con referentes o estándares externos a la región, mientras que en la prueba del Laboratorio se esperaba llegar a acuerdos regionales sobre medidas deseables de logro.

Una última diferencia que vale la pena resaltar es que, mientras las técnicas de muestreo y los instrumentos de las pruebas de la IEA y de la OCDE son apreciados por su rigor y validez, el Laboratorio enfrentó serios cuestionamientos técnicos por parte de los coordinadores nacionales. Por otro lado, la participación en las pruebas de ambas agencias ha favorecido la reflexión sobre los contenidos curriculares de cada país, en la medida en que las comparaciones revelan diferencias importantes en los enfoques epistemológicos y en la distribución de contenidos a lo largo del currículo de educación básica. Sin embargo, esa reflexión se ha dado principalmente entre técnicos y expertos de las unidades de medición, pero con escasa o ninguna participación de los profesionales de las áreas de desarrollo curricular. Desde el punto de vista de la política curricular, ello implica que la medición internacional en cada país tiene escasa vinculación con los procesos de implementación de los programas curriculares que, según se expresa, se desean mejorar mediante las evaluaciones.

Sin embargo, los aspectos más débiles de los proyectos de evaluación, según se observa en la experiencia de la mayor parte de los países latinoamericanos, han sido el de la **interpretación de los resultados** y su posterior **diseminación**. Además de las insuficiencias en el planeamiento estratégico, la escasa difusión de resultados estuvo marcada por la falta de comunicación y coordinación entre los equipos técnicos de medición y los de desarrollo curricular en cada país. Estos dos factores –falta de una estrategia de difusión y falta de trabajo conjunto entre actores clave– determinaron que se redujeran las posibilidades de impacto en el mejoramiento de la implementación de reformas y en la reformulación de objetivos curriculares.

Durante la prueba del Laboratorio, la debilidad en la interpretación y diseminación se debió,

principalmente, a la falta de una estrategia preconcebida, así como a la falta de un marco conceptual sólido y de estándares que permitieran una adecuada interpretación sustantiva de los resultados, a la vez que el segundo informe oficial, que se esperaba tuviera más impacto de política, llegó a los ministerios demasiado tarde, cuando las administraciones nacionales habían cambiado. Cuba fue, de acuerdo a las entrevistas, una excepción en este sentido, dado que allí los buenos resultados obtenidos fueron ampliamente difundidos desde la Presidencia.

En cuanto a los efectos de los resultados de pruebas de la IEA, es evidente que las desfavorables posiciones de América Latina en los *rankings* internacionales han causado un efecto de *shock* en algunas audiencias, especialmente por la forma de reporte que ha utilizado la prensa. En la actualidad, los países que han decidido continuar participando en pruebas de la IEA y de la OCDE están conscientes de la necesidad de “educar” a la prensa y otras audiencias influyentes, de volverlas más especializadas, de manera que puedan difundir informaciones más relevantes que la temida tabla de posiciones. Los reportes que redactan las oficinas de evaluación o currículo de los ministerios, por su parte, se consideran particularmente útiles cuando pueden tener impacto en el ámbito local, escolar y pedagógico y cuando contribuyen a desarrollar capacidades de análisis e interpretación de datos. El caso de Estados Unidos durante la prueba TIMSS fue singular, por el alto involucramiento de diversos grupos de interés y la consulta sistemática a actores del sector, así como también por la fuerte estrategia de difusión utilizada.

Con respecto al **impacto** de la participación en pruebas internacionales en las decisiones sobre política educativa, de repercusiones políticas, o de investigaciones realizadas sobre los datos, nuevamente se hace evidente la diferencia entre la experiencia de América Latina y lo ocurrido en Estados Unidos. En nuestra región, tanto la evaluación del Laboratorio como las pruebas de la IEA han tenido un impacto político y de política relativamente bajo. Las excepciones, según nuestros datos, son Cuba, que tomó algunas decisiones concretas de políticas sobre la base de los resultados del Laboratorio, y Chile, que ha debido enfrentar el impacto sobre la opinión pública de los resultados en las pruebas de la IEA. En el caso del Laboratorio se han realizado algunas investigaciones posteriores, en los países

y por parte de la OREALC misma, pero hasta el momento casi no han sido difundidas.

Por otra parte, las implicancias curriculares de las pruebas internacionales en América Latina, han sido casi nulas puesto que no se registra evidencia de que los resultados hayan alertado a los profesionales e investigadores de la educación sobre la pertinencia o adecuación de los contenidos curriculares nacionales. La excepción, hasta donde pudo verificar este estudio, es también Cuba dado que allí la experiencia de evaluación comparada permitió alertar a los representantes nacionales sobre la necesidad de introducir cambios en la secuenciación de algunos conceptos matemáticos, así como sobre el enfoque teórico con que se plantean los contenidos en el área de lenguaje.

En cuanto a la participación en pruebas de la IEA, la experiencia ha repercutido casi exclusivamente en el desarrollo técnico de las unidades de medición, lo cual es significativo, pero no se reportan medidas importantes de política curricular, o educativa en general, en función de los resultados.

Con respecto a PISA, una ventaja anticipada del proyecto que puede repercutir positivamente en el medio educativo es que los instrumentos mismos de medición, serán propiedad de los países participantes, y se podrán utilizar en aplicaciones nuevas e independientes para medir otras muestras y así favorecer investigaciones más profundas en el ámbito nacional.

En Estados Unidos, donde existen más de cien años de trayectoria en medición y evaluación educacional, la experiencia ha sido muy diferente, particularmente con relación a las pruebas TIMSS y TIMSS-R de la IEA. La creación de consorcios educativos, el involucramiento de organizaciones sindicales docentes y asociaciones profesionales de las diferentes disciplinas académicas, así como las múltiples líneas de investigación surgidas a partir de los resultados, confirman que el impacto de la participación en pruebas internacionales depende no solamente del potencial informativo de esas pruebas, sino también de las estrategias de sensibilización previas a la difusión final de los resultados⁵.

Estas diferencias entre las experiencias de Estados Unidos y América Latina confirman que **la ausencia de un diálogo informado entre los funcionarios y usuarios de la información sistemática**, resulta en un escaso impacto de las evaluaciones en las decisiones sobre políticas educativas, y sobre políticas curriculares en particular. (Ver recuadro 1).

Adicionalmente, dado que no se monitorean los resultados sobre las pocas decisiones que se toman en función de las informaciones de estudios comparados, es difícil conocer si se ha avanzado o no en la dirección deseada. Pero hasta el momento la contribución de los resultados de pruebas internacionales a nuevos ciclos de política curricular, o de mejoramiento de los existentes, es mínima y en muchos casos inexistente. En otras palabras, según indican los datos de las pruebas internacionales, es muy probable que las iniciativas y decisiones centrales de política curricular que se tomen en el futuro no consideren la evidencia sobre los resultados de políticas y programas anteriores.

V. RECOMENDACIONES PARA FUTURAS OPORTUNIDADES DE PARTICIPACIÓN EN PRUEBAS INTERNACIONALES

La evidencia recogida en este trabajo, si bien no se puede considerar representativa de toda la región, ha permitido arribar a algunas conclusiones iniciales sobre la conveniencia de la participación de países latinoamericanos en futuras pruebas internacionales, así como de las condiciones en que esa participación debería darse.

1. Dado que muchos de los países de la región ya han participado en alguna prueba internacional con anterioridad, se recomienda utilizar la información de esas pruebas para generar opinión y sensibilizar a audiencias de interés antes de iniciar nuevos programas de reforma o mejoramiento curricular, y antes de adscribirse a un nuevo proyecto de evaluación comparada internacional. Esta medida tendría el valor adicional de permitir el pilotaje de posibles estrategias de difusión de los resultados.

Una forma viable de sensibilizar a los representantes de la prensa, a las organizaciones sindicales y a los profesionales de la

⁵ Vale recordar que para cuando la IEA liberó los resultados de TIMSS, diferentes audiencias y medios de información ya estaban expectantes y deseosos de conocer los resultados.

educación es invitarlos a participar en las discusiones iniciales del proyecto, así como en la discusión sobre resultados anteriores, sobre la importancia de iniciar nuevas evaluaciones y sobre el vínculo concreto que se percibe entre las evaluaciones y los contenidos curriculares. Es importante que los coordinadores en el ámbito nacional calculen con anticipación, y en colaboración con otras oficinas ministeriales, el costo material y el tiempo de trabajo que requerirá la fase de discusión y sensibilización, para que esta sea un aspecto formal y viable del proyecto de evaluación.

Durante las discusiones, y mediante la información que se haga circular sobre resultados anteriores, se sugiere alertar sobre los posibles bajos resultados comparativos del país, pero simultáneamente ejemplificar la manera en que la próxima prueba comparada puede informar sobre aspectos específicos de desarrollo o implementación curricular. Paralelamente, esta fase puede comenzar a “educar” a las audiencias convocadas en algunos de los tecnicismos propios de la psicometría y de la sociometría, para que comiencen a desarrollar algunas capacidades de interpretación y análisis de datos más allá de los *rankings*.

Si el proyecto de evaluación está liderado por una institución académica o no gubernamental, es especialmente importante reunir a las autoridades ministeriales de medición y currículo, no solo para obtener el aval oficial, sino para establecer una especie de “contrato” sobre las necesidades de información percibidas por el ministerio, u organización equivalente, y el tipo de información que puede derivarse de la evaluación comparada.

2. Al momento de decidir sobre el tipo de prueba internacional que se va aplicar, cada país debería considerar los siguientes **aspectos técnicos y contextuales**, que afectarán la mayor o menor utilidad de las experiencias de evaluación comparada:

a. El **momento histórico** de las reformas curriculares en curso, la duración prevista del gobierno en funciones, y el grado de afianzamiento del sistema nacional de medición de la calidad educativa.

b. Las **necesidades reales de la información** que se va a recoger, y las posibilidades concretas de aprovecharla, de acuerdo a los momentos de desarrollo e implementación curricular del país y de acuerdo a las dificultades percibidas o comprobadas anteriormente en esos procesos.

c. Las **características de las pruebas**, es decir, el marco conceptual utilizado, los dominios disciplinarios o cognitivos evaluados, y los factores escolares y socioeconómicos relevados con relación al rendimiento. Si la prueba es de cobertura regional y se presenta como un proyecto de elaboración conjunta entre los países participantes, es importante definir un **marco conceptual** que considere el contexto de política educativa y curricular de la región y permita una interpretación contextualizada de los resultados del estudio. Un estudio de esta naturaleza debería representar también una oportunidad para definir con los técnicos de los demás países un conjunto de **estándares** mínimos de contenido para realizar una revisión de los currículos nacionales y comparar, mediante la evaluación, los alcances relativos de los currículos prescritos, implementados y logrados.

d. Las **oportunidades de participación efectiva** que el proyecto ofrece con relación a la construcción o adecuación de instrumentos de medición, o a la inclusión de “opciones nacionales” de evaluación; y la capacidad técnica de los coordinadores nacionales para hacer efectiva y viable esa participación.

e. La **utilidad y pertinencia del informe** de resultados utilizado por la agencia evaluadora, de acuerdo al tipo de audiencias que se espera informar.

3. Durante la etapa de implementación del proyecto se recomienda motivar y asegurar el vínculo y la **comunicación permanente** entre las unidades de medición y los equipos de desarrollo curricular en los niveles central y regional en cada país. Si parte de estos procesos es eminentemente técnica y requiere poca intervención de personal de currículo, se recomienda mantenerlos

informados permanentemente sobre los avances del proyecto y sobre las fechas en que serán convocados para discutir y analizar los resultados del estudio. Simultáneamente, debería mantenerse una comunicación permanente con los representantes de la prensa y de las asociaciones sindicales y profesionales convocadas inicialmente (recomendación 1).

El mismo mecanismo debería operar entre los coordinadores y los actores gubernamentales y no gubernamentales si el operativo de evaluación estuviera en manos de una institución académica independiente. Estos mecanismos de comunicación permanente son una forma de mantener vivo el “contrato” entre los técnicos y los consumidores de información que más pueden influir en la toma de decisiones para cambios de política y práctica curricular. Como existen altas posibilidades de que los usuarios gubernamentales (ministeriales) de la información cambien súbitamente, es más importante aún reforzar los canales de comunicación permanente con la prensa y con representantes de instituciones sindicales, académicas, profesionales y de la sociedad civil.

4. Hacia el momento en que se reciben los datos ya procesados de la agencia evaluadora, y de acuerdo a las recomendaciones anteriores, los posibles usuarios de la información deberían estar informados de las fechas en que serán convocados a **analizar e interpretar** los datos de la prueba. Aquí se sugiere realizar primero una sesión de análisis y discusión entre los técnicos de evaluación y los representantes de desarrollo curricular, para planificar en conjunto una agenda de discusión organizada para el momento de intercambio con representantes de la prensa, del mundo académico y de organizaciones no gubernamentales. Las discusiones con estos actores deberían conducir a una serie de recomendaciones puntuales sobre el tipo de informes que sería deseable difundir, concretamente con relación al formato, al tipo de información que se debe incluir, a las audiencias prioritarias y secundarias que se quiere alcanzar, al número sugerido de publicaciones y a los diferentes objetivos que los informes deberían perseguir (por ejemplo, si deben brindar in-

sumos de información para medidas de política o para decisiones de tipo curricular y pedagógico en el nivel local y escolar).

Según indica la evidencia, la difusión de la información por medio de informes no es suficiente para lograr un impacto real en las políticas y prácticas educativas. Por ello, es necesario organizar eventos locales con amplia participación de directores de escuela y docentes, en los cuales se explique cómo interpretar los datos y se discutan las diferentes maneras de mejorar el desarrollo curricular en las escuelas y su implementación en las aulas. Este es también el momento de **revisar el contexto de política curricular**: en qué momento de implantación de los programas curriculares se encuentran el país y sus regiones; cuáles son las dificultades que evidencian las pruebas nacionales, los informes de supervisión regionales, las investigaciones académicas y la prueba internacional misma para decidir qué tipo de mejoras (si no cuáles exactamente) se pueden recomendar a las autoridades nacionales y subnacionales de educación.

En este sentido, el análisis de los datos debería conducir a preguntas clave sobre la forma en que otros países logran mejores rendimientos estudiantiles y sobre las condiciones bajo las cuales esos logros son posibles. Vale recordar que una de las ventajas de las pruebas internacionales es que pueden revelar información sobre los efectos diferenciales de algunas variables sobre el logro de aprendizajes que no es posible detectar en un estudio intranacional. También es importante recordar que las asociaciones entre variables dependientes e independientes que muestran los estudios internacionales deberían tomarse con cautela y ser analizadas dentro del contexto particular de cada país.

5. Se recomienda hacer un seguimiento del tipo de decisiones de política y práctica educativa, y curricular en particular, tomadas en función de los resultados difundidos. Una vez identificadas, se sugiere comisionar o alentar a instituciones académicas el estudio de los resultados que sobre el aprendizaje de los estudiantes han tenido, o no, esas decisiones y cambios de prácticas.

La información recogida mediante estos estudios debería canalizarse y difundirse adecuadamente para ofrecer evidencia más concreta para la toma de decisiones sobre nuevas políticas y programas curriculares, así como para la decisión sobre nuevas evaluaciones internacionales y el tipo de estrategias que son más efectivas para la difusión y uso de sus resultados.

6. Una vez que se ha participado en una experiencia internacional de evaluación, y si la información obtenida de esa experiencia está a la altura de las expectativas de los coordinadores nacionales y de las autoridades políticas del sector, se aconseja continuar participando en nuevas mediciones, especialmente en las siguientes aplicaciones de la misma prueba. Esto permitiría la medición del progreso académico de las poblaciones evaluadas y, además, que el público y los usuarios de la información se acostumbren a contar con esta periódicamente, y de esta manera maximizar su uso técnico y político.
7. Se recomienda la participación de unidades subnacionales de gestión educativa (por ejemplo, estados, departamentos o grupos de escuelas asociadas). Si la participación no puede ser financiada por esas unidades, como probablemente ocurriría en América

Latina, la coordinación central de la evaluación (ministerios u otras agencias a cargo) debería considerar la posibilidad de conseguir recursos para promover esa participación. Es deseable, también, que la muestra nacional, en vez de estar conformada por alumnos de todo el país, sea construida con casos de un departamento o estado que presente la suficiente heterogeneidad demográfica como para asegurar variabilidad en los resultados y, por ende, la posibilidad de analizar estadísticamente su relación con las variables predictivas del logro. Ello permitiría diseñar más fácilmente estrategias de difusión y uso de los datos en un contexto más reducido, para luego replicar esas estrategias en el ámbito nacional.

Esperamos con este trabajo haber contribuido a la comprensión de las condiciones técnicas y políticas bajo las cuales los países de América Latina deberían participar en pruebas internacionales. Es nuestro deseo que los resultados de este informe de investigación alcancen y se difundan en los espacios de decisión sobre evaluación educativa en la región, así como en el ámbito académico, y así coadyuvar a una mejor utilización de los recursos disponibles para la medición de rendimiento académico y, lo que es más importante aun, para su impacto en la calidad de los aprendizajes.

RECUADRO 1

DIÁLOGO ENTRE INVESTIGADORES, EVALUADORES Y USUARIOS DE LA INFORMACIÓN

Uno de los aspectos que representa el mayor obstáculo para el uso efectivo de la evaluación internacional en el mejoramiento de la calidad de los aprendizajes, según se ha detectado en este estudio, es la ausencia de diálogo entre investigadores o evaluadores y los usuarios potenciales de la información.

Una explicación al respecto se puede encontrar en Reimers (1999), quien sostiene que en las altas esferas de administración educativa en muchos países del mundo, la mayor parte de las decisiones que se toman para incrementar las oportunidades de aprendizaje se realizan intuitivamente, especulando con la articulación de algunos de los elementos del sistema (currículo, capacitación, supervisión, etc.), en vez de tomarse como resultado del análisis cuidadoso de la evidencia que muestran las investigaciones y los análisis de política. El autor argumenta que esto se debe a la forma en que el conocimiento derivado de la investigación es acumulado y procesado, y a que no siempre se reconoce la diferencia entre la investigación que pretende simplemente explicar la realidad, y la investigación de política que, además, busca describir cómo podría cambiarse esa realidad, que se asume es lo que los funcionarios necesitan saber.

Esa función propositiva de la investigación debería darse en forma de **diálogo entre los investigadores, funcionarios y otros usuarios de la información**, de tal manera que se logre un acuerdo entre las necesidades de información y las posibilidades reales de los investigadores de proveerla.

El modelo racional de toma de decisiones, que asume que los funcionarios seleccionarán opciones de política sobre la base de un análisis informado, está siendo cuestionado. Se reconoce que la mayor parte de los problemas no están, ni pueden ser, perfectamente definidos y que los funcionarios no siempre establecen las prioridades de acción racionalmente sino en función de variables de oportunidad política y de poder. Por ello, para lograr decisiones informadas, es preciso ejercer cierto grado de persuasión en los decisores de políticas mediante un proceso de negociación, y no de diseño unilateral, del proceso de reforma. Ese proceso debe ser interdisciplinario, propositivo –y no solamente descriptivo–, y debe tener en cuenta que el análisis se debe realizar y comunicar dentro de los límites de tiempo que imponen los ciclos de política.

Adicionalmente, Reimers propone el uso de un “mediador de conocimiento” entre el mundo de la investigación y el mundo de la política. La evidencia recogida para este estudio muestra que, en ciertas ocasiones, esa mediación se logra a través de la prensa y de otras instituciones académicas y no gubernamentales formalmente convocadas durante el proceso de evaluación internacional. Sostiene Reimers que “el problema del uso de la información no puede ser considerado recién después de concluida la investigación o el análisis [...] la investigación debe comenzar con la consideración de las necesidades del cliente si se intenta evitar el riesgo de que los resultados del análisis resulten irrelevantes para los funcionarios”.

BIBLIOGRAFÍA

- ATKIN, Myron y Paul BLACK (1997). *Policy perils of international comparisons: The TIMSS case*. Phi Delta Kappan, 79,1, 22-28.
- BALL, Stephen (1998). "Big policies/small world: An introduction to international perspectives in education policy". *Comparative Education*, 34, 2, 119-130.
- BEATON, Albert *et al* (1999). *The benefits and limitations of international educational achievement studies*. International Institute for Educational Planning, International Academy of Education. Unesco, París.
- BEMPECHAT, Janine *et al.* (Borrador, 2000). "Cultural-cognitive issues in academic achievement: New directions for cross-national research". *Board on International Comparative Studies in Education (BICSE)*, National Research Council.
- BLOOM, Benjamin (1974). "Implications of the IEA studies for curriculum and instruction". *School Review*, 82, 3, 413-435.
- BURSTEIN, L. (1989). "Conceptual considerations in instructionally sensitive assessment". Documento presentado en American Educational Research Annual Meeting, San Francisco.
- BURSTEIN, L. *et al.* (1992). "Education indicators", en M. Atkin (Ed.), *Encyclopedia of educational research* (6ª edición). Macmillan, Nueva York.
- CARTER, D.S.G y M.H. O'NEILL (1995). *International perspectives on educational reform and policy implementation*. Falmer, Brighton.
- GRADE (2001). "Mejoramiento de los procesos de desarrollo curricular y de materiales educativos". Informe final de la consultoría N° 010-2000-ED/MECEP para el Ministerio de Educación del Perú.
- FARRELL, Joseph (1979). *The necessity of comparisons in the study of education: The salience of science and the problem of comparability*. Presidential address, annual meeting of CIES, Ciudad de México. Reimpreso en *Comparative Education Review*, 23, 1. 1978.
- FERRER, Guillermo (1999). "Aspectos del currículum prescrito en América Latina: Revisión de tendencias contemporáneas en currículum, indicadores de logro, estándares y otros instrumentos". GRADE-PREAL. <http://www.grade.org.pe/gtee-preal/docs/Ferrer.pdf>
- FORSTER, Margaret (2000). *A policy maker's guide to international achievement studies*. ACER Press, Australia.
- GREANEY, Vincent y Thomas KELLAGHAN (1996). *Monitoring the learning outcomes of education systems*. Banco Mundial, Washington DC.
- HUSÉN, Torsten (1994) "Educational research and policy-making", en T. Husén y T. Postlethwaite (Eds), *International Encyclopaedia of Education*. Pergamon Press, Oxford.
- HUSÉN, Torsten (1987). "Policy impact of IEA research". *Comparative Education Review*, 31, 1, 29-46.
- HUSÉN, T. y A. TUIJNMAN (1994). "Monitoring standards in education: Why and how it came about", en A.Tuijnman y T. Postlethwaite (Eds), *Monitoring the Standards of Education*. Pergamon Press, Oxford.
- JONES, P.W (1998). "Globalisation and internationalism: Democratic prospects for world education". *Comparative Education*, 34, 133-155.
- KELLAGHAN, Thomas (1996). "IEA studies and educational policy". *Assessment in Education*, 3, 2, 133-160.
- KOHN, Melvin (1987). "Cross-national research as an analytic strategy" (American Sociological Association, 1987 Presidential Address). *American Sociological Review*, 52, diciembre, 713-731.
- LEVIN, Benjamin (2001). "Conceptualizing the process of education reform from an international perspective". *Education Policy Analysis Archives*, 9, 13. <http://epaa.asu.edu/epaa/v9n13.html>
- LYOTARD, J.F (1984). *The postmodern condition: A report on knowledge*. Manchester University Press, Manchester.

MACNAB, Donald (2000). "Forces of change in mathematics education: the case of TIMSS". Educational Policy Analysis Archives 8,15.

MCKNIGHT, Curtis (1987) *The underachieving curriculum: Assessing U.S. school mathematics from an international perspective*. A National Report on the Second International Mathematics Study.

MISLEVY, Robert (1995). "What can we learn from international assessments?" *Educational Evaluation and Policy Analysis*, invierno 1995, 17, 4, 419-437.

National Commission on Excellence in Education (1983). *A nation at risk*. Government Printing Office, Washington DC.

PLATT, W.J (1975). "Policy making and international studies in educational evaluation", en A.C. Purves y D.U. Levine (Eds.) *Educational policy and international assessment*. McCutchen, Berkeley, CA.

REIMERS, Fernando y Noel MC GINN (1997). *Informed Dialogue: Using Research to Shape Education Policy Around the World*. Praeger, Nueva York y Londres..

SCHMIDT, William y Curtis MCKNIGHT (1995). "Surveying educational opportunity in mathematics and science: An international perspective", *Educational Evaluation and Policy Analysis*, otoño 1995, 17, 3, 337-353.

SMITH, Adam (1776). *An inquiry into the nature and causes of the wealth of nations*. Dublin, Whitestone, 1776. 3 v. 22 cm. The Library of Congress, EEUU.

THEISEN, Gary *et al* (1983). "The underachievement of cross-national studies of achievement". *Comparative Education Review*, 27, 1, 46-68.

TORNEY-PURTA, Judith (1990). "International comparative research in education: Its role in educational improvement in the US". *Educational Researcher*, (volumen, S/D) 32-35.

TORRES, Rosa María (1999). "*Reformadores y docentes: El cambio educativo atrapado entre dos lógicas*", Documento de discusión. IIPE Unesco, Buenos Aires.

WIRT, Frederick (?) "*Comparing educational policies: Theories, units of analysis, and research strategies*". Reimpreso en *Comparative Education Review*, vol.24, 2, junio1980.

WOLF, R (1979). "*Sampling*". Boletín # 4: *Secondary Study of Mathematics*. Second International Mathematics Study, Urbana, IL.

ENTIDADES ENTREVISTADAS

(Por razones de confidencialidad, solamente se reportan los nombres de las agencias o unidades ministeriales a la que pertenecen los 25 entrevistados de este estudio)

Argentina

- IDECE (Instituto para el Desarrollo de la Calidad Educativa). Ministerio de Educación.
- Ex Subsecretaría de Evaluación de la Calidad Educativa. Ministerio de Educación.

Chile

- Programa Formación Inicial Docente. División Educación Superior.
- Unidad de Curriculum y Evaluación/ SIMCE. Ministerio de Educación.

Colombia

- ICFES (Instituto Colombiano para el Fomento de la Educación Superior)
- Universidad del Valle. Ciudad de Cali.

Cuba

- Instituto Central de Ciencias Pedagógicas.

Ecuador

- APRENDO. EB/PRODEC. Sistema semiautónomo oficial de medición de la calidad de la educación.

Estados Unidos

- Universidad de Wisconsin-Madison.
- Universidad Estadual de Michigan/ Universidad Estadual de Nueva York, Albany.
- Federación Estadounidense de Profesores.
- Departamento de Educación. (American Federation of Teachers. Educational Issues Department.)
- Consorcio SMART (Science and Mathematics Required for Tomorrow).

México

- Subsecretaría de Planeación y Coordinación Institucional, Secretaría de Educación Pública.

Organización de Estados Iberoamericanos (OEI)

- Programa Evaluación de la Calidad de la Educación.

Perú

- Radio Programas del Perú.
- Unidad de Medición de la Calidad. Ministerio de Educación.
- UNESCO-Perú.

Unesco-OREALC

- Laboratorio Latinoamericano para la Evaluación de la Calidad de la Educación (LLECE).

Uruguay

Unidad de Medición de Resultados Educativos (UMRE) de la Administración Nacional de Educación Pública.

SIGLAS Y ABREVIACIONES

AFT: American Federation of Teacher, 'Federación Estadounidense de Maestros'.

ALL: Adult Literacy and Lifeskills, 'Prueba para Adultos sobre Alfabetización y Destrezas para la Vida'.

BID: Banco Interamericano de Desarrollo

ETS: Educational Testing Services, 'Servicios de Evaluación Educacional'.

ICFES: Instituto Colombiano para el Fomento de la Educación Superior.

IDECE: Instituto para el Desarrollo de la Calidad Educativa.

IEA: International Association for the Evaluation of Educational Achievement, 'Asociación Internacional para la Evaluación del Logro Académico'.

IRT: Item Response Theory, 'Teoría de Respuesta al Item'.

LLECE: Laboratorio Latinoamericano para la Evaluación de la Calidad de la Educación.

NCES: National Council for Education Statistics, 'Consejo Nacional de Estadística de la Educación'.

NCTM: National Council of Teachers of Mathematics, 'Consejo Nacional de Profesores de Matemáticas'.

NSF: National Science Foundation, 'Fundación Nacional de Ciencias'.

OCDE: Organización para la Cooperación y el Desarrollo Económico.

OEI: Organización de Estados Iberoamericanos.

ONG: Organización no-gubernamental.

OREALC (Unesco): Oficina Regional para América Latina y el Caribe.

PIRLS: Progress in International Reading Literacy Study, 'Estudio Internacional sobre los Progresos en Alfabetización y Habilidades de Lectura'.

PISA: Programme for International Student Assessment, 'Programa Internacional de Evaluación de Estudiantes'.

PNUD: Programa de las Naciones Unidas para el Desarrollo.

PROMEDLAC: Comité Regional Intergubernamental del Proyecto Principal de Educación en América Latina y el Caribe.

REDUC: Red Latinoamericana de Información y Documentación en Educación.

SIMS: Second International Mathematics Study, 'Segundo Estudio Internacional de Matemáticas'.

SACMEQ: Southern Africa Consortium for Monitoring Educational Quality, 'Consortio de África del Sur para la Evaluación de la Calidad Educativa'.

SEP: Secretaría de Educación Pública.

SIMCE: Sistema de Medición de la Calidad Educativa.

SMART: Science and Mathematics Achievement Required for Tomorrow.

TIMSS: Third International Mathematics and Science Study, Tercer Estudio Internacional en Matemáticas y Ciencias'.

TIMSS-R: Repetición de TIMSS.

UMRE: Unidad de Medición de Resultados Educativos.

WEI: World Education Indicators, 'Indicadores Mundiales de Educación'.