



ISBN: 978-607-02-2345-7

Universidad Nacional Autónoma de México

Instituto de Investigaciones
sobre la Universidad y la Educación

www.iisue.unam.mx/libros

Catalina Inclán Espinosa (2011)

“La prueba PISA en el contexto de América Latina”
en *La prueba PISA 2006. Un análisis de su visión
sobre la ciencia,*

Ángel Díaz-Barriga (coord.),

IIISUE-UNAM, México, pp. 23-52.

Esta obra se encuentra bajo una licencia Creative Commons
Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional
(CC BY-NC-ND 4.0)

LA PRUEBA PISA EN EL CONTEXTO DE AMÉRICA LATINA

*Catalina Inclán Espinosa**

ANTECEDENTES SOBRE SISTEMAS INTERNACIONALES DE EVALUACIÓN

Los primeros estudios internacionales para medir el rendimiento educativo surgen con las agencias internacionales de evaluación, que tienen por finalidad establecer comparaciones sobre el rendimiento escolar en varios países y comprobar de esta manera tanto la efectividad de los sistemas como el aprendizaje de los alumnos.

Como pionera tenemos a la International Association for the Evaluation of Educational Achievement (Asociación Internacional para la Evaluación del Logro Educativo) (IEA) con sede en Ámsterdam, cuyos orígenes datan de los años cincuenta; se trata de un consorcio cooperativo independiente de institutos y agencias de investigación en más de 45 países, que desde hace 40 años se ha ocupado de realizar mediciones y estudios de contexto orientados a conocer el nivel y condiciones en que se da el aprendizaje de las matemáticas y las ciencias, en los estudiantes del nivel básico y medio superior.¹

* Investigadora del ISSUE-UNAM.

1 "Los orígenes de la IEA se remontan a 1958 cuando un grupo de académicos, psicólogos educativos, sociólogos y expertos en psicometría se reunieron en el Instituto de la UNESCO para la educación en Hamburgo a fin de descubrir los problemas de la escuela y la evaluación de los estudiantes. Argumentaron que una evaluación efectiva requiere [...] el examen de los resultados educativos [...] Este estudio, conocido como Pilot Twelve-Country Study, se llevó a cabo en 1959-1962 con muestras de estudiantes de 13 años de edad en 12 países [...] en cinco áreas: matemáticas, comprensión de lectura, geografía, ciencia y capacidad no verbal."

Su misión, al realizar estudios comparados sobre las políticas y prácticas educativas de los países participantes, es influir en el aprendizaje dentro y a través de los sistemas de evaluación, y no sólo medir el rendimiento escolar en cuanto a resultados de aprendizaje, sino examinar los efectos de los programas, la organización de las escuelas y las aulas, la relación entre rendimiento y actitudes, los efectos de las prácticas en algunas asignaturas, el tiempo pasado en clase y la enseñanza de la redacción en el idioma materno, así como diferentes prácticas educativas y los resultados obtenidos por grupos especiales. Entre sus objetivos de evaluación destaca un nuevo concepto de rendimiento, que incluye la valoración no sólo de los resultados, sino de los factores a los que obedecen dichos resultados:² “El rendimiento educativo es valorado tomando en consideración diversas variables de contexto y de proceso, con el objetivo de buscar explicaciones e interpretaciones a las diferencias de rendimiento y no limitarse solamente a la construcción de indicadores o a la difusión de tablas de clasificación de los países participantes.”³

El primer estudio internacional de la IEA⁴ fue La Enseñanza Elemental y Secundaria de las Matemáticas en los países, efectuado entre 1959 y 1966. En 1964 se realizó el Primer Estudio Internacional sobre Matemáticas (FIMS). Las ciencias se evaluaron por primera vez en los años 1970-1971 con el First International Science Study (Primer Estudio Internacional sobre Ciencias). A éste le siguieron el Study of Civic Education (Estudio sobre Educación Cívica) en

En 1964 realiza también en esos países el First International Mathematics Study (FIMS) y no es hasta 1967 que obtiene su personalidad jurídica. A Forshay, *Brief history of IEA*, <http://www.iea.nl/brief_history_of_iea.html>.

- 2 Cabe señalar que este objetivo todavía no es algo consolidado, en virtud de que a pesar de los esfuerzos de las agencias internacionales de evaluación por incluir estudios de contexto para conocer los factores asociados al rendimiento escolar, la forma como hasta ahora se han presentado los resultados del rendimiento escolar y de los cuestionarios de contexto no contribuye a una comprensión del porqué de ellos. Creemos que para que este objetivo se pueda cumplir satisfactoriamente se hace necesario pasar de los puntajes, porcentajes, índices y gráficas, a una significación pedagógica de lo ocurrido en dichas evaluaciones.
- 3 Véase A. Tiana, *Cooperación internacional en evaluación de la educación en América Latina y el Caribe*, p. 14.
- 4 La IEA fue la primera organización internacional que utilizó las mismas pruebas objetivas en más de un país.

1971;⁵ en 1980-1982 el Second International Mathematics Study (SIMS), aplicado a 224 países, y en 1984 el Second International Science Study (SISS). Entre 1989 y 1992 se realizó el Computers in Education (estudio realizado sobre tecnologías de la información y la comunicación); finalmente, en 1990-1991 se aplicó el Reading Literacy Study (RLS) (primer estudio sobre lectura).⁶

LA EVALUACIÓN EN AMÉRICA LATINA

En América Latina prácticamente todos los países crearon sistemas nacionales de evaluación de la calidad; así, se instalaron mecanismos regionales de medición de resultados y diversos países se incorporaron a las mediciones internacionales.

Existen varios trabajos que hacen un análisis sobre qué ha significado en materia de política educativa esta “fiebre evaluativa”. Lo que salta a la vista es que la mayor parte de estas evaluaciones indican que la calidad de la educación no ha mejorado, que no hay cambios significativos, salvo el relativo al crecimiento de la matrícula (lo que implica que se ha incorporado a alumnos de sectores con mayores dificultades socioculturales, antes excluidos del sistema educativo). Esta situación reclama un análisis sobre la eficacia de las reformas educativas que se han impulsado en la región desde la década de los noventa.

Es necesario tener en cuenta que los sistemas de medición han convertido los resultados en motivo de debate público: el impacto que tiene la difusión de los resultados de pruebas nacionales o internacionales en los medios de comunicación se ha convertido en

5 La descripción del estudio realizado en 28 países se encuentra en: IEA, *Estudio de educación cívica*, <http://www2.rz.hu-berlin.de/empir_bf/iea_s.html>. Un resumen de los resultados, en el documento: J. Torney-Purta y L. Orozco, *Ciudadanía y educación cívica en 28 países: resumen de los resultados del estudio comparado de la IEA*, <http://www.wam.umd.edu/~iea/exec_summ/ExecutiveSummSpanish.htm>.

6 Otros estudios de evaluación internacional: *International Assessment of Educational Progress* (IAEP-1); se trata de un estudio sobre matemáticas y ciencias realizado por el Board on International Comparative Studies in Education en 1988, el cual se repitió con el IAEP-2 en 1991.

una de las principales dimensiones que debe manejar toda gestión de política pública en educación. En estos momentos no es posible pensar la evaluación exclusivamente desde el punto de vista técnico-pedagógico, sino que es evidente la pertinencia de una perspectiva analítica mucho más amplia y compleja.

Es conveniente reconocer que los docentes percibieron estos dispositivos más como una presión o una amenaza que como un insumo para el mejoramiento de las estrategias de enseñanza-aprendizaje. Por otro lado, la información de los resultados no se transformó en insumo para el mejoramiento de las políticas, los datos no se elaboraron en las escuelas y el impacto sobre las estrategias de enseñanza fue muy débil.

En esta medida, vale la pena preguntarse sobre la finalidad que tiene la evaluación, es decir sobre su para qué, a fin de intentar establecer un sentido de la acción que se pretende construir. La evaluación en este momento parece tener múltiples finalidades, tales como mejorar la eficiencia, compensar diferencias sociales, comprobar niveles de segmentación, juzgar el sistema educativo y ponerlo en evidencia y legitimar políticas. Clarificar el sentido de la evaluación coadyuvaría a entender qué se pretende con su puesta en marcha en la región.

ALGUNAS CONSIDERACIONES SOBRE LOS SISTEMAS DE EVALUACIÓN EN AMÉRICA LATINA

La experiencia de evaluación en la región a partir de los años noventa permite realizar una serie de consideraciones que no pueden pasarse de largo al aplicar sistemas de medición y evaluación de resultados de aprendizaje; entre ellas podemos destacar las siguientes:

Fines y propósitos de la evaluación: ¿para qué evaluar?

Los propósitos de la evaluación y los usos de los resultados deben ser materia de debate desde el inicio de un proceso de establecimiento de un sistema de evaluación educacional. Las bondades del enfoque,

metodología, instrumentos e impacto de cada examen a gran escala tienen que establecerse en función del propósito para el que están concebidos y afectan directamente el uso de sus resultados.

Es conveniente clarificar el objetivo y el sujeto de la evaluación. Lo que significa diferenciar si la pretensión es evaluar el funcionamiento del sistema educativo, evaluar el desempeño de las instituciones escolares o evaluar a los actores de la educación de manera individual, sean docentes o estudiantes.

Es prioritario reconocer que la elaboración de una prueba a gran escala no necesariamente ofrece información sobre todos estos objetivos; de ahí que en ocasiones la premura de su diseño y la forma de difundir sus resultados induzca a usos equivocados de la información que se obtiene en ellas. La ausencia de una visión de largo plazo o la necesidad de improvisar medidas de “emergencia” debilita la legitimidad de estos sistemas. En América Latina la experiencia de esta situación es claramente indicativa de este problema; así, por ejemplo, en la República Dominicana existe una unidad de medición interna, y las pruebas de opción múltiple están concebidas con referencia tanto a una norma como a un criterio. Por su parte, en Chile se han realizado pruebas de medición del rendimiento escolar desde 1982 a través del Sistema de Medición de la Calidad Educativa (SIMCE), que es un organismo autónomo del Ministerio de Educación, cuyos instrumentos están elaborados para evaluar contenidos mínimos obligatorios y estándares nacionales para todos los grados, a través de instrumentos cerrados y abiertos. En este caso se elaboran informes para los docentes con la pretensión de que tengan una comprensión de los resultados de tales pruebas, mientras que en el ámbito de las políticas se establecen programas compensatorios para aquellos sectores que muestran un rezago académico.

Entidades que realizarán la medición: ¿quién va a evaluar?

Sobre este tema también existen experiencias muy diferentes en la región, de acuerdo con el grado de independencia que tengan los organismos evaluadores de la autoridad educativa; tal es el caso de

Colombia, con el Instituto Colombiano para el Fomento de la Educación Superior, si bien, este organismo tiene un carácter semiautónomo. Por el contrario existen organismos evaluadores que se encuentran estrechamente vinculados con los ministerios de educación, como es el caso de Ecuador, Guatemala, Bolivia o Perú (Unidad de Medición de la Calidad y nueva ley de educación que establece el Instituto Peruano de Evaluación, Acreditación y Certificación Educativa). También se pueden identificar los llamados institutos, que tienen una mayor autonomía administrativa y técnica en general, como en Brasil el Instituto Nacional de Estudios e Investigaciones Educativas o en México el Instituto Nacional de Evaluación de la Educación.

Relación entre aprendizaje, currículo y estándares: ¿qué evaluar?

Los instrumentos de medición deben estar referidos a las expectativas o metas de aprendizaje que cada país se haya propuesto en sus estructuras curriculares y puedan ser reconocidas por todos los involucrados en el sistema: docentes, directores, supervisores y expertos. Esto constituye un marco de interpretación básico que permitirá comprender los datos arrojados por las evaluaciones.

Es habitual encontrar que los planes de estudios nacionales ofrecen definiciones muy genéricas y, en estricto sentido, poco claras de lo que se espera que los alumnos puedan realizar; es decir, las expectativas que se tienen sobre su aprendizaje. En algunos países de la región (Uruguay, Colombia, Ecuador, Chile) se ha abierto un análisis entre expertos o entre expertos y docentes sobre los contenidos de enseñanza de cada ciclo o nivel educativo, e incluso en algunos lugares esta discusión se realiza sobre lo que en este momento se denomina estándares de desempeño o desarrollo de competencias. En el fondo, este análisis orienta con mayor claridad a los docentes sobre lo que se espera de su trabajo y a los especialistas en elaboración de pruebas a gran escala, sobre los puntos relevantes en torno a los cuáles hay que construir las preguntas del examen. Es lamentable obser-

var que en México el grupo de expertos curriculares y especialistas en elaboración de libros de texto, esto es, los analistas del contenido, pertenecen a la Subsecretaría de Educación Básica, mientras que los responsables de elaborar las pruebas a gran escala se encuentran en un organismo autónomo o en otra sección de la Secretaría de Educación Pública como la Coordinación de Planeación. Ello ocasiona un importante desencuentro: las pruebas miden los contenidos que consideren relevantes, los especialistas curriculares hacen sus definiciones y orientaciones en sus oficinas de la subsecretaría y los profesores tratan, en el mejor de los casos, de interpretar ello, cuando no se limitan a trabajar con lo que, a su juicio, emana de los libros de texto.

Instrumentos: ¿cómo evaluar?

Existen dos formas de aplicar las pruebas a gran escala: muestrales y censales. La política educativa de cada país en América Latina emplea criterios diversos para decidir cuándo aplicar una u otra estrategia. Las pruebas censales son empleadas para calificar, por así decirlo, un tramo de la formación; el caso más claro lo constituyen las pruebas de salida o egreso que se aplican en Chile, República Dominicana o Costa Rica, al término de la enseñanza media. Podemos reconocer que la prueba IDANIS (Instrumento de Diagnóstico para Alumnos de Nuevo Ingreso a la Educación Secundaria) que resuelven los alumnos en el 6° grado de primaria en México tiene indirectamente estos rasgos.

Por su parte, las pruebas muestrales buscan reconocer una tendencia en el logro de aprendizajes escolares. La construcción de dicha muestra requiere la delimitación de una serie de variables sociales, económicas y culturales que permitan construir una geografía nacional de los diferentes tipos de escuelas que se tiene, sea por su integración: urbanas, semiurbanas, rurales; sea por el grado de desarrollo: integradas-marginadas; sea por el grupo social que las frecuenta: ciudadinas, rurales, indígenas o de migrantes. En el caso de Bolivia, Guatemala y Perú se observa la aplicación de pruebas a gran escala de acuerdo con el índice de discriminación, validación técnica,

o adecuación lingüística; mientras que en el caso de Uruguay sobresale la incorporación de factores asociados al rendimiento académico, contextual.

México, en su incremento casi exponencial de pruebas a gran escala aplica en todo el sistema educativo ambos modelos. Una censal: ENLACE, la más conocida, aunque también se podría considerar la prueba estatal que desarrolla simultáneamente y por separado cada entidad federativa, o la prueba IDANIS para alumnos del sexto grado de primaria; y una muestral: EXCALE, que elabora el INEE, y otras que podrían tener un rasgo similar aunque con una variable acotada, como puede ser la prueba Factor de Aprovechamiento Escolar, para los alumnos cuyos docentes están en el Programa de Pago al Mérito llamado Carrera Magisterial. En un exceso, un alumno de sexto grado en el país puede resolver de manera simultánea hasta cinco exámenes a gran escala elaborados con criterios técnicos, de aprendizaje y de contenido diferentes.

Reporte de resultados: ¿cómo informar?

La forma en que se reportan los resultados es importante si se busca que éstos sean comprendidos por quienes pueden mejorar la educación. En este punto es necesario tener presente que no son las mismas necesidades de información las que tiene un especialista en test, un tomador de decisiones, los diversos responsables del trabajo en las escuelas (supervisores, directores y grupos de maestros), o la sociedad en general.

Esta falta de discriminación sobre quién es el usuario de los resultados de un examen ocasiona que el contenido de los informes de evaluación se distorsione. En general, los resultados son retomados por los medios de comunicación para enfatizar las deficiencias de los aprendizajes que muestran los alumnos, y, por su parte, a los políticos les permiten hacer declaraciones sobre el tema. Desde otro ángulo la sociedad, además de impactarse por los resultados exige (como una exigencia hetérea) que se mejore la educación y, aliada a los medios de comunicación, empieza a realizar clasificaciones

(rankeos) de alumnos y escuelas y aun directivos del sistema; pero, sobre todo, los docentes, se perciben en el banquillo de los acusados, y se encuentran sin una clara orientación sobre cuáles son las acciones o las estrategias que tienen que realizar para mejorar el trabajo escolar.

Los reportes suelen carecer de una distinción sobre el nivel económico y sociocultural de la población estudiantil, y de las condiciones de equipamiento de las escuelas, información que evitaría conclusiones injustas. Sin embargo, la dimensión más ausente en los reportes de evaluación es la didáctico-pedagógica, que finalmente es la que permitiría orientar las acciones de supervisores, directivos y docentes para establecer una estrategia de mejora, y para involucrar a los padres de familia e incluso a los medios de comunicación, en particular los medios masivos, en la responsabilidad social (y no sólo escolar) de mejorar el aprendizaje de los alumnos.

Con gran esfuerzo, en algunos casos se elaboran recomendaciones en torno a lo que puede hacer una escuela en particular; tal es el caso de algunos reportes chilenos, brasileños (Minas Gerais) y panameños. En Uruguay, por su parte, se fomenta el establecimiento de talleres para analizar los informes de evaluación.

TEMAS QUE SE DISCUTEN EN MATERIA DE EVALUACIÓN DE LOS APRENDIZAJES

Cuando uno se pregunta sobre lo que aportan las evaluaciones a gran escala, en comparación con sus costos, parecen mayores los costos y el beneficio político para las autoridades educativas nacionales que los beneficios pedagógicos.

Las evaluaciones se han vuelto un hecho político: los gobiernos justifican su responsabilidad social frente a la educación al presentar resultados de evaluación, pero, en estricto sentido, no se consideran responsables de ellos ni toman en cuenta los errores del pasado, referidos a las decisiones políticas y pedagógicas (planes de estudio, programas de formación docente y proletarización de la profesión docente, así como el papel que le han dejado a los medios de comunicación masiva para conformar un aprendizaje social lejano y

contradictorio al que busca la escuela). Los políticos y autoridades educativas nacionales utilizan los resultados de la evaluación exclusivamente para exhibir las deficiencias de la escuela y de los docentes: éstos deben dar los mejores resultados con independencia de los planes de estudio, de la infraestructura escolar, o de las condiciones socioculturales de los alumnos. El mundo educativo ha perdido el manejo exclusivo de esta herramienta; ello significa que, como indican ciertas opiniones, es muy importante realizar análisis de los resultados de la evaluación, pues, de lo contrario, esa información sólo apuntala las decisiones de política educativa. Pero hay que decirlo: si no queda claro para qué se está evaluando, los sistemas educativos dejarán de trabajar para mejorar la calidad y la equidad, y pasarán a trabajar en el mejoramiento de los resultados de la evaluación. Tal es el caso de todas las ediciones de libros que se han generado en América Latina para entrenar a los alumnos en la resolución de este tipo de exámenes.

EL USO DE LA INFORMACIÓN EN LAS PRUEBAS A GRAN ESCALA

Las pruebas a gran escala se emplean en mayor número en los sistemas educativos de América Latina. A diferencia de lo que acontecía hace 20 o más años, hoy se pueden encontrar en internet múltiples reportes nacionales o internacionales que dan cuenta del desempeño de los estudiantes en estos exámenes. La primera impresión que se lleva el interesado es que en general los estudiantes de educación básica en América Latina obtienen resultados deficientes en las pruebas que se aplican para determinar el grado en que han logrado aprendizajes en ámbitos como los de matemáticas y español. Los informes que dan cuenta de las diferentes pruebas que se aplican en la región permiten observar claramente esta situación.

Las sociedades de América Latina han sido objeto de un bombardeo de información, fundamentalmente la que aparece en diversos encabezados de los medios de comunicación, en los que se presenta, de una manera un tanto alarmante, la deficiencia de resultados que muestran los sistemas educativos.

Si tomamos como ejemplo algunos titulares de diarios de la región, se encuentran expresiones como las siguientes: “A pesar de la reforma educativa cayó el rendimiento de los alumnos”;⁷ “Alumnos argentinos, casi los últimos en leer y escribir”;⁸ “Alumnos brasileños se encuentran en el último lugar en el ranking de educación”;⁹ “SIMCE revela un gran desastre en la educación. Cifras revelan mínimo avance tras la reforma de 1996”;¹⁰ “Matemáticas e inglés el ‘coco’ de los estudiantes”;¹¹ “Escuelas bolivarianas reportan bajo rendimiento académico”;¹² “Estudiantes no pueden con destrezas básicas. Bajo rendimiento reportó el Sistema Nacional de Medición del Logro”;¹³ “Alumnos de tercer grado, apenas lograron el 55% de rendimiento”;¹⁴ “Perú, penúltimo lugar entre 15 países de América Latina en nivel educativo”.¹⁵

La sociedad mexicana no ha estado exenta de esta situación. A principios de los años noventa se publicaba el libro *La catástrofe silenciosa*,¹⁶ en donde se sostiene la tesis de una ineficiencia educativa responsable del bajo aprovechamiento escolar. En este libro apareció por primera ocasión un término que se generalizaría en la primera década del siglo XXI: “México: un país de reprobados”.¹⁷

7 *La Nación*, Buenos Aires, 21 de mayo de 1998.

8 *La Nación*, Buenos Aires, 9 de abril de 2003.

9 *Folha on Line*, <<http://www.folha.com.br>>, 4 de diciembre de 2001.

10 *La Tercera*, Santiago de Chile, 1 de enero de 2001.

11 *El País*, Bogotá, 2 de mayo de 2002.

12 *El Universal*, Caracas, 7 de enero de 2007.

13 *El Expreso*, Guayaquil, 23 de agosto de 1999.

14 *ABC*, Montevideo, 1 de abril de 1998.

15 *La República*, Lima, 25 de mayo de 2000.

16 Guevara, Gilberto, *La catástrofe silenciosa*, 1992.

17 La expresión *país de reprobados* se refería a la siguiente situación: ineficiencia escolar que ha originado un rezago acumulado de 25 millones de adultos sin estudios primarios; discrepancia e incongruencias entre los planes de estudio de los diversos niveles del sistema educativo; normatividad y legislación poco claras y eficientes; irrelevancia de los contenidos educativos; mecanismos de evaluación deficientes y que por lo regular pocas veces reflejan de manera objetiva el aprendizaje adquirido; desproporción entre matrícula y crecimiento demográfico, y oferta y demanda de maestros; insuficiente financiamiento de la educación; frustrados intentos de descentralización.

Es de llamar la atención que en dicha obra el vocablo *reprobados* se utilizaba reconociendo múltiples causas del sistema educativo en ello, mientras que en el artículo publicado por esos años el autor se refiere exclusivamente al rendimiento de los estudiantes.¹⁸

Todas estas experiencias han contribuido a aceptar “una cultura de la evaluación”, aunque su riesgo es que la evaluación se convierta en un objetivo por sí mismo, con todo y no ser posible reducir el hecho educativo a un indicador medible. Ahora bien, en una decidida visión contraria vale la pena interrogar de frente esta “cultura de la evaluación” y sus presupuestos ideológicos, y preguntarse de una vez si no hay misiones de la escuela que no pueden restringirse a ser evaluadas.

Las preocupaciones de las últimas décadas en relación con la educación son: acceder a ella, permanecer, terminar el año, completar un programa de estudio y un libro de texto, pasar el examen, aprobar un ciclo, recibir un diploma o certificado.¹⁹ Cuando se crea una amplia cantidad de indicadores cuantitativos se corre el riesgo de que las escuelas funcionen sólo para satisfacerlos, en detrimento de la tarea central de éstas: la formación.

Podremos saber mucho de los resultados de los alumnos a través del rendimiento escolar e ignorar al mismo tiempo todo acerca de sus procesos y la incidencia de los aprendizajes fuera del ámbito escolar (en la familia, la comunidad, la participación cívica). La evaluación, por más que lo ha intentado, tampoco puede incorporar todos los factores extraescolares que entran en contacto diariamente entre docentes y alumnos: la familia, los amigos, la calle, el barrio, la comunidad, el juego, el periódico, la televisión, la computadora, la cancha, etcétera.

No hay que olvidar que toda empresa que mide, escoge lo que quiere medir, selecciona aquello que privilegia. A este respecto habrá que preguntarse si la escuela debe enseñar a responder y no a preguntar; si la escuela debe enseñar a resolver problemas matemáticos pero no problemas de la vida diaria; si la escuela debe educar para

18 G. Guevara Niebla, “México: un país de reprobados”, *Nexos*, núm. 162, junio de 1991.

19 R. M. Torres, *12 tesis para el cambio educativo*, <<http://www.entreculturas.org/node/390>>.

cuando esos niños y jóvenes sean adultos (en otro contexto y condiciones), o para que entiendan el presente y su condición actual.

EL COSTO DE LA EVALUACIÓN EN LA REGIÓN: ¿MAYOR GASTO EN EVALUACIÓN, MENOR GASTO EN FORMACIÓN DOCENTE?

Hay poca información sobre el costo de la evaluación en la región y no existen análisis sobre el porcentaje que representan estos gastos con respecto al total que cada país invierte en el sector educativo; mucho menos se realiza un análisis sobre el gasto en la evaluación y en los sistemas de formación y perfeccionamiento docente (cuadro 1). Los escasos reportes que existen al respecto invitan a preguntarse hasta dónde América Latina no está realizando un gasto excesivo en sus sistemas de evaluación sólo para reconocer algo que las pruebas indican: hay deficiencias en el sistema educativo. En este punto se imponen las preguntas de si no es el momento de apostar, aun desde el punto de vista financiero, a mejorar el trabajo de aprendizaje en el aula, disminuyendo los recursos que se emplean para evaluar; y si la evaluación debe seguir excluyendo de su construcción, de sus problemas, de su producción y de sus resultados a los docentes, impidiendo con ello que incorporen su experiencia en la elaboración de los criterios, de los indicadores, de la discusión de sus resultados.

Ninguna de las evaluaciones nacionales o internacionales ha tenido a bien dirigirse a los docentes, traducir sus resultados en implicaciones para su tarea, reconocer y no enjuiciar su trabajo, orientar posibles relaciones pedagógicas, discutir caminos didácticos y establecer estrategias de aprendizaje que le permitan al docente (el actor más cercano a los procesos de aprendizaje) comprender dónde y cómo tienen lugar los aprendizajes, u ofrecerle información que le permita inventar constantemente fórmulas pedagógicas capaces de hacer trabajar juntos a alumnos heterogéneos. Sobre todo a aquellos estudiantes que para los criterios de la evaluación no logran el estándar o salen de la norma, por lo cual son evidenciados o expulsados.

CUADRO 1

Compendio de los costos de la evaluación en América Latina

País	Tipo de evaluación	Año	Asignatura	Grado	Costo total (US\$ ppp)*	Núm. estudiantes evaluados	Costo total por estudiante (US\$ ppp)	Costo de evaluación en % total presupuesto por nivel educativo**	
								Primaria	Secundaria
Chile	Universal	2004	Lengua, matemáticas, Ciencias naturales, ciencias sociales	8°	4 472 000	300 000	15		0.17
Colombia	Internacional	2001	Lectura	4°	599 000	5 131	117	0.01	
	Universal	2002	Lengua, matemáticas	5°, 9°	2 548 000	1 030 626	2	0.02	0.02
	Universal	2003	Civismo, ciencias	5°, 9°	5 497 000	1 034 049	5	0.04	0.05
	Universal	2004	Examen de acceso a la educación superior	11°	9 418 000	540 716	17		0.17
	Muestra	2004	Lengua, matemáticas, ciencias naturales	5°, 9°	1 363 000	96 242	14	0.01	0.01
Honduras	Muestra	2002	Lengua, matemáticas, ciencias naturales	3°, 6°	1 953 000	42 572 000	46	0.28	
	Muestra	2004	Lengua, matemáticas, ciencias naturales	3°, 6°	2 292 000	45 657 000	50	0.33	
Perú	Internacional	2000	Matemáticas, lectura, ciencias	Mayores de 15 años en 7° o superior	480 000	5 190	92		0.04
	Muestra	2001	Matemáticas, lectura y redacción	4°, 6°, 10°	1 757 000	34 000	52	0.06	0.05
	Muestra	2004	Civismo, matemáticas, lectura y redacción	2°, 6°, 9°, 11°	4 885 000	70 000	70	0.15	0.13

Fuente: Laurence Wolff, *The costs of student assessments in Latin America*, p. 17.

CUADRO 1
(continuación)

País	Tipo de evaluación	Año	Asignatura	Grado	Costo total (US\$ ppp)*	Núm. estudiantes evaluados	Costo total por estudiante (US\$ ppp)	Costo de evaluación en % total presupuesto por nivel educativo**	
								Primaria	Secundaria
Uruguay	Internacional	2003	Matemáticas, resolución de problemas, lectura, ciencias	Mayores de 15 años en 7° o superior	311 000	5 797	54		0.08
	Muestra	2001	Desarrollo cognitivo y afectivo, lengua, matemáticas	Preescolar, 1°, 2°	599 000	5 131	117	0.01	
	Muestra	2002	Lengua, matemáticas	6°	111 000	9 171	12	0.03	
	Muestra	2003	Lengua, matemáticas, ciencias naturales, ciencias sociales	12°	266 000	12 993	20		0.07

* ppp= paridad del poder adquisitivo. El total no incluye los gastos generales.

** El presupuesto total de la educación por nivel corresponde a los valores de 2000. Cuando la prueba se lleva a cabo tanto en la enseñanza primaria como en la secundaria se presupone que los costos se dividen equitativamente entre ambos niveles. Los gastos generales no están incluidos. Para aclaraciones sobre las diferencias halladas en esta tabla entre costos totales absolutos y costos unitarios, véase la sección "La variabilidad de los costos de las evaluaciones", más adelante.

Clave: *Universal*: evaluaciones aplicadas a todos los alumnos de los grados seleccionados. *Muestra*: evaluación nacional por muestras, es decir las que se aplican a grupos de alumnos seleccionados al azar, de manera que se pueden extrapolar conclusiones respecto al universo no evaluado. *Internacional*: participación en pruebas internacionales de evaluación.

Una experiencia de evaluación en el contexto latinoamericano es la que desarrolló el Laboratorio Latinoamericano de Evaluación de la Calidad Educativa (LLECE),²⁰ sistema de evaluación creado en 1994 especialmente para los países de la zona por la Oficina Regional de Educación para América Latina y el Caribe (OREALC)²¹ de la UNESCO; concebido, más que como un estudio propiamente comparativo, como una Red de los Sistemas Nacionales de Evaluación surgida a partir de la década de los noventa en el marco de los objetivos y recomendaciones del Proyecto Principal de Educación en América Latina y el Caribe en el periodo 1993-1996,²² representa la parte operativa y concreta de uno de los programas estratégicos de la OREALC denominado Evaluación de la Calidad de los Aprendizajes.²³

- 20 Los orígenes de la prueba LLECE se le atribuyen a la UNESCO cuando en el año de 1989 este organismo lleva a cabo a través de pruebas de rendimiento y encuestas un estudio comparativo exploratorio en siete países de la región, cuyos resultados dejaron de manifiesto las deficiencias existentes en los sistemas educativos. En la exposición de estos resultados, los representantes regionales plantearon a la UNESCO-OREALC la necesidad de contar con un mecanismo de evaluación que permitiera la obtención de información sistemática a través de una organización estable. Es así como surge la idea de crear el Laboratorio Latinoamericano para la Evaluación de la Calidad de la Educación, después más comúnmente conocido como "Laboratorio", a partir del cual comienza a gestarse la idea de realizar una prueba comparativa propia para la región latinoamericana. Otro de los acontecimientos que por su impacto originó la creación de este "Laboratorio" fue el del Proyecto Principal de Educación en América Latina y el Caribe (PROMEDLAC), iniciado en 1980 para terminar en el año 2000, como resultado del consenso de los representantes gubernamentales de los países de la región sobre la necesidad de un esfuerzo renovado, intenso y sostenido para atender, teniendo como horizonte el año 2000, carencias y necesidades educativas fundamentales no satisfechas e indispensables para enfrentar para el desarrollo de los países, respondiendo a objetivos compartidos por todos y mediante acciones en que el esfuerzo propio de cada país se beneficie de la cooperación horizontal, subregional, regional e internacional (OREALC/UNESCO, *Necesidades básicas de aprendizaje*). De ahí que en 1993 los ministros reunidos en Chile discutieran la necesidad de que la UNESCO impulsara el desarrollo de la evaluación y la creación de estándares en la región.
- 21 Los programas promovidos por OREALC/UNESCO, obedecen a los principios del Informe Delors para el siglo XXI. J. Delors, *La educación encierra un tesoro*.
- 22 Para tener una idea más clara de los alcances de este proyecto se puede consultar OREALC/UNESCO, *Situación educativa de América Latina y el Caribe 1980-2000*.
- 23 La puesta en marcha de sistemas de evaluación en la región está estrechamente vinculada con la prioridad política de mejorar la calidad a través de la evaluación, lo que comenzó a darse primero en países más desarrollados y luego se hizo extensivo a América Latina con

Este Primer Estudio Internacional Comparativo sobre Lenguaje, Matemáticas y Factores Asociados para Alumnos de Tercero y Cuarto Grado de la Educación Básica realizado por el LLECE se aplicó a un número aproximado de 48 688 niños y a 41 077 familiares, 3 675 docentes y 1 387 directores de trece países de la región,²⁴ cuyos resultados fueron publicados en 1998.²⁵

Entre los objetivos del LLECE está el generar estándares²⁶ regionales de aprendizajes, medir el estado de los niveles de aprendizaje logrados, tener información respecto a la situación del aprendizaje escolar y las variables que en éste inciden, obtener información sobre la calidad y estado de las variables, y experimentar el enfoque innovador de algunos métodos pedagógicos; para elaborar una comparación del rendimiento académico y analizar los factores asociados a éste, así como fortalecer los sistemas nacionales de medición y evaluación por vías prácticas y generar una vía de investigación a partir de los resultados obtenidos.²⁷ Lo que se busca es establecer niveles de cali-

los llamados sistemas nacionales de evaluación surgidos en la última década del siglo XX. "La evaluación es concebida como un poderoso instrumento al servicio de la mejora cualitativa de la educación, gracias al conocimiento valorativo que proporciona sobre la realidad" (A. Tiana, *op. cit.*, p. 4).

24 Véase A. Tiana, *op. cit.*, pp. 7-8. Los países que participaron en este primer estudio fueron: Argentina, Bolivia, Brasil, Colombia, Cuba, Chile, Costa Rica, Honduras, México, Paraguay, Perú, República Dominicana y Venezuela.

25 Los resultados de este primer estudio que presenta los logros obtenidos por once de los trece países participantes (el Informe no incluye los resultados de Perú y Costa Rica por circunstancias técnicas y de autorización) se publicaron en 1998 (Casassus *et al.*) y se presenta una visión comparada de los logros obtenidos en materia de aprendizaje en países con características culturales semejantes.

26 El Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación, define a los estándares como "Construcciones (constructos teóricos) de referencia...útiles para llevar adelante acciones en algún ámbito determinado" (Casassus, "Marcos conceptuales...", p. 3).

27 Los sistemas nacionales de evaluación de la calidad surgen a partir de la década de los noventa, como parte importante de una política educativa que cobra vigor ante los niveles de expansión experimentados por los sistemas educativos en la década de los ochenta en América Latina; pues una vez alcanzada la cobertura deseada en cuanto a oportunidades de acceso, lo que interesa ahora son los contenidos y procesos de la experiencia escolar, para así mejorar la calidad de la educación en cuanto a procesos y variables que inciden en los niveles y resultados de aprendizaje obtenidos por los estudiantes. Existen varios estudios y documentos que se han dedicado al análisis y difusión de estos sistemas de evaluación, como son los de Horn *et al.*, *Sistemas de medición y evaluación educacional...*; Rojas y Esquivel, *Sistemas de medición del*

dad y eficiencia en los sistemas educativos en aras de determinar los conocimientos y destrezas requeridos para promover la participación activa de jóvenes en el mercado de trabajo y apreciar los efectos de la enseñanza sobre el aprendizaje.

La intención del LLECE es incorporar a los planes y programas de estudio los niveles de logro e indicadores de calidad del aprendizaje, mediante el establecimiento de estándares que intentan servir como construcciones teóricas de referencia para la acción y para la rendición de cuentas.²⁸ Éstos comprenden cuatro dimensiones: lo prescrito, objetivos educativos que constituyen el elemento básico; lo deseable, que alude a elementos de excelencia en cuanto a lo que se espera de la educación; lo observable, es decir, lo que efectivamente se logra, y lo factible, referente a las condiciones para su realización.²⁹

En 1997 surge la necesidad de aplicar pruebas de lenguaje y matemáticas para realizar el Primer Estudio Internacional Comparativo sobre Lenguaje, Matemáticas y Factores Asociados en Tercero y Cuarto Grado. Con estos exámenes el LLECE buscaba conocer lo que aprenden los alumnos; en qué ciclo o grado escolar lo han aprendido y qué tipo de competencias han desarrollado. Para lo cual se generaron instrumentos de medición y evaluación consistentes en pruebas de matemáticas y lengua, y cuestionarios de contexto, con la intención de conocer las aptitudes, capacidades, competencias y destrezas

logro académico...; Schwartzman, "Las pruebas de medición del desempeño escolar..."; y Arregui, Sistema de determinación de metas... Actualmente los sistemas nacionales de evaluación se encuentran en una etapa de revisión y mejora en cuanto a su función, calidad técnica, uso y difusión de los resultados. El Programa de Promoción de la Reforma Educativa en América Latina y el Caribe (PREAL), en conjunto con Grupo GRADE (Grupo de Análisis para el Desarrollo) tienen varios documentos al respecto, véase J. Esquivel, ¿Cómo evaluar los aprendizajes en América Latina?, 2001, y P. Ravela, ¿Cómo presentan sus resultados los sistemas nacionales de evaluación...?, 2001.

28 *Rendición de cuentas* es concepto emanado del mundo de trabajo que destaca, una vez más, la inserción de conceptos laborales en el ámbito educativo y el enfoque pragmático, que encierra para la educación y la evaluación el establecimiento de estándares. "Si no existiera esta relación con la acción, no tendría mucho sentido elaborar estándares en la educación, pues la observación de la calidad en la educación, conlleva a la necesidad de formular acciones para mejorarla... al no existir estándares nacionales, la sociedad no tiene los elementos de información necesarios para exigir *accountability* al sistema" (Cassasus, "Marcos conceptuales...", p. 6).

29 *Loc. cit.*

de los alumnos;³⁰ conceptos que se constituyen como estándares básicos, relacionados con el aprendizaje, que refieren lo que se espera de un sistema educativo.³¹

LA PRUEBA PISA Y SU IMPACTO EN AMÉRICA LATINA

A partir del año 2000 la Organización para la Cooperación y Desarrollo Económicos (OCDE), a través de su Programme for International Student Assessment (PISA), dio inicio a la aplicación de pruebas a gran escala para jóvenes de 15 años en varios países. El estudio se concibió en 1997 y se puso en marcha al año siguiente. Fue aplicado por primera vez en el 2000 a un número aproximado de 250 000 estudiantes, principalmente de los países miembros de esta organización; sin embargo, es importante destacar que algunos países no miembros solicitaron que dicho examen también se aplicase a sus alumnos.³²

En el año 2000 se evaluaron estudiantes de 43 países; en 2003, 41, y en 2006, 57 países (30 miembros de la OCDE y 27 asociados). La prueba se aplica cada tres años, y en cada ocasión se pone especial atención en un área distinta; en el 2000 fue lectura, en 2003 matemáticas y en 2006 ciencias, aunque en las tres ocasiones se aplican preguntas de las tres áreas.

Los principales intereses a los que responde son: obtener indicadores sobre el rendimiento en las áreas de lectura, matemáticas y ciencias de los estudiantes de 15 años; contar con elementos que

30 Los documentos del Laboratorio refieren la aptitud como disposición innata relacionada con el rendimiento que permite diferenciar a los individuos con un mismo nivel mental. La capacidad se refiere a poder realizar un acto físico o mental, puede ser innata o desarrollada mediante el aprendizaje. Y la competencia alude a la formación o preparación de un individuo para intervenir de un modo eficaz en un proceso de contexto (Cassasus, "Marcos conceptuales...", p. 9).

31 Según documentos del Laboratorio, "cuando se quiere generar estándares, es inevitable referirse a lo que se espera que los alumnos aprendan y que los profesores enseñen. Es necesario establecer de forma clara y pública, cuáles son las competencias conceptuales y prácticas que se espera que alcancen los alumnos" (Cassasus, "Marcos conceptuales...", p. 6).

32 Entre los primeros países no miembros de la OCDE que decidieron incorporarse están Brasil, Latvia, Liechtenstein y Federación Rusa.

permitan influir en la definición de las políticas educativas e identificar y difundir los factores que en cada sistema educativo favorecen el desarrollo de competencias en la perspectiva de habilidades y destrezas para la vida que desde mediados de los años noventa se debate en Europa.

De ahí que no se puede decir que sea una prueba centrada en conocimientos únicamente. Por ejemplo, desde su versión de 2000 el informe PISA contenía una afirmación parecida a la siguiente: “si sólo tienes los conocimientos de las materias que se examinan no vas a poder resolver esta prueba, si no los tienes, tampoco”.³³ Así, queda de manifiesto que no es una prueba que mida la capacidad de retención y aplicación rutinaria de un conocimiento. El interés está centrado en conocer en qué medida la educación logra que estudiantes de 15 años, que están por concluir un nivel educativo e ingresar a un nivel de educación superior, utilicen la información que han aprendido para resolver problemas concretos y reales de la vida cotidiana.

Cabe destacar que además esta prueba cuenta con un marco teórico para cada una de las disciplinas evaluadas. Lo que en palabras de Díaz Barriga sería la *teoría de contenido*.³⁴ Lo que quiere decir que PISA cuenta con marcos de evaluación: en 2000, con una serie de conceptualizaciones sobre la lectura; en 2003, sobre las matemáticas, y en 2006, sobre las ciencias.

Esto revela diferencias significativas entre estudios internacionales de evaluación como PISA y algunas pruebas aplicadas en México como el Examen Nacional de Ingreso EXANI-I, del Centro Nacional de Evaluación CENEVAL. Estudios como el de la OCDE organizan su contenido conforme a la definición que en dichas pruebas se hace. Esto es, la prueba PISA, antes de presentar los resultados, ofrece una conceptualización acerca de lo que se podría denominar una teoría de la lectura (2000), de las matemáticas (2003) o de las ciencias (2006). Por su parte, las pruebas aplicadas en el ámbito nacional a estudiantes regulares de 15 años que se encuentran en tercero de secundaria (como el EXANI-I o el más reciente examen que aplica

33 Véase OCDE, *Conocimientos y aptitudes para la vida. Resultados de PISA 2000*.

34 Véase A. Díaz Barriga, “ENLACE-EXCALE: tensiones y retos”.

la Secretaría de Educación Pública, denominado Evaluación Nacional de Logro Académico en Centros Escolares, mejor conocido como ENLACE) son pruebas alineadas al currículo; esto es, carecen de una teoría del contenido. Su contenido y estructura se da con base en los planes y programas de estudios y en la taxonomía de Bloom (1971).³⁵

Hay, sin embargo, otras diferencias: el concepto de contenido de evaluación en estudios como PISA está enfocado a comprobar y comparar el rendimiento escolar en cuanto a comprensión y aplicación de los conocimientos. Las pruebas realizadas en México continúan con contenidos centrados en la reproducción de los temas del plan de estudios. El tipo de ítem de los exámenes internacionales evalúa capacidades cognitivas para la aplicación del conocimiento a situaciones de la vida cotidiana, mediante el uso de procedimientos. Los ítems de los exámenes nacionales requieren de operaciones básicas de pensamiento centradas en la identificación y el recuerdo, para la realización de ejercicios rutinarios. Todo ello da cuenta de que las pruebas mexicanas, en general, le conceden mayor importancia a la función enciclopédica de la educación. Y en el mejor de los casos algunas de sus preguntas incluyen temas un tanto utópicos para la realidad que vive el joven en México.

Además, las pruebas internacionales incorporan tendencias recientes en materia de elaboración de exámenes como la llamada teoría de respuesta al ítem. Son pruebas centradas en desempeños y contextualizadas. Las pruebas de la experiencia mexicana conservan los modelos de la llamada teoría clásica. Cabe mencionar que en las últimas versiones de la prueba ENLACE ya se incorpora este enfoque, pero lo apresurado de su elaboración hace dudar de que reúnan los rangos estadísticos que se exigen a una prueba de tal magnitud.

Tenemos entonces que tanto contenidos curriculares básicos, como estándares y preparación de los jóvenes para la vida adulta, son los conceptos base para la organización del contenido de evaluación en las pruebas internacionales como PISA; además de que se abordan los factores sociales, escolares y familiares que influyen en los resultados obtenidos. Lo que para México sigue siendo un reto.

35 B. Bloom, *Taxonomía de los objetivos de la educación*.

El examen de PISA evalúa no sólo conocimientos, sino también capacidades y/o aptitudes para la aplicación de esos conocimientos a situaciones cotidianas, plantea un concepto de contenido de corte funcional y procedimental, y presenta pruebas que descartan el manejo enciclopédico del contenido curricular para dar paso a la evaluación de contenidos básicos y procesos superiores de pensamiento, en contextos reales; se trata de aspectos que en México todavía no han sido generalizados ni trabajados a profundidad.

Además, PISA está estructurado con base en un concepto de competencia, considerada como habilidad y/o aptitud para la aplicación de un conocimiento y, aunque carece de una definición precisa, presenta una idea más funcional de aprendizaje y evaluación.

Por las características conceptuales de los contenidos en los estudios de evaluación internacional, se puede decir que éstos retoman los planteamientos curriculares de John Dewey referentes a contenidos inmersos en la experiencia o la práctica y de Franklin Bobbit, quien desde 1918 planteaba la necesidad de articular el currículum al mundo laboral.

LA PRUEBA PISA Y LA REALIDAD SOCIOCULTURAL Y EDUCATIVA DE MÉXICO Y AMÉRICA LATINA

Estudios sobre el contenido de la prueba PISA permiten identificar algunas dificultades. Por ejemplo, en el caso de matemáticas existe un reactivo que hace referencia a la unidad monetaria del ZEDS, lo que presenta una dificultad adicional que puede alterar el resultado, y hace pensar si los errores de los alumnos pudieran deberse a un problema de lenguaje (no lograron traducir *zeds* por unidad monetaria, o bien sólo lograron resolver el problema en su nivel matemático). En los informes de PISA no encontramos datos sobre este importante problema técnico.

El concepto de contenido que subyace en la prueba PISA obliga a reparar en el tipo de contenidos trabajados en la escuela básica y en los factores asociados en materia de desarrollo curricular; porque los estudios de evaluación internacional sobre el aprovechamiento

escolar no sólo contemplan lo relacionado con el rendimiento en términos de resultados, sino que además ofrecen una evaluación de los factores asociados al rendimiento y desempeño escolar que incluye lo referente a procesos y situaciones que se dan en la práctica educativa.

El informe de la prueba PISA cuenta con una referencia en cuanto al nivel esperado de respuestas de los estudiantes. Este nivel lo obtiene mediante una combinación de índices de calidad de vida (socioeconómicos) e inversión estatal en educación. Lo que no se ha difundido es que los alumnos mexicanos que obtuvieron en promedio respuestas en el nivel 2 de desempeño en ciencias, se encuentran dentro del rango del nivel esperado. Esto lleva a pensar que la prueba PISA demanda realizar no únicamente ajustes pedagógicos en las escuelas, en los planes de estudios y en la formación de maestros, sino que requiere de la conformación de una nueva estructura social y cultural en el país.

Además de problemas en materia de aprendizaje, los resultados de PISA revelan las grandes desigualdades sociales que vive México. Los informes presentan que los estudiantes con menores rendimientos fueron los de Oaxaca, Chiapas, Tabasco y Guerrero, mientras que los estudiantes que obtuvieron puntajes más altos corresponden al Distrito Federal, Aguascalientes, Coahuila, Nuevo León y Querétaro.³⁶

Como he señalado, la prueba PISA se aplica cada tres años, razón por la cual no se tienen variaciones en los indicadores, ya que el lapso entre cada prueba es corto. No hay duda, pues, de que en México hay un exceso de evaluación a un costo altísimo y no necesariamente con una mejora educativa.

Hoy los estudiantes de secundaria tienen la prueba IDANIS al ingreso, la ENLACE en el tercer grado, la prueba PISA cada tres años, el EXANI I (o su equivalente para el ingreso al bachillerato); ante lo cual, después de aplicar una prueba, lo único que ocurre es que se toman escasas decisiones para mejorar la educación y se difunden en la prensa los resultados. La reflexión sobre el concepto de examen

36 Véase M. A. Díaz Gutiérrez *et al.*, *PISA 2006 en México*.

de la prueba PISA y lo que esto implica en materia de elaboración de exámenes es algo que no se considera en los informes nacionales sobre PISA.

Las tendencias en materia de elaboración de exámenes a gran escala que se han encontrado, tanto en el plano internacional como nacional, presentan un importante cambio en la concepción de examen que los especialistas en elaboración de exámenes a nivel nacional deberían considerar: presentan conceptos de contenido, aprendizaje y evaluación diferentes; el contenido que se evalúa está centrado más en habilidades para la aplicación del conocimiento, que en la reproducción de temas curriculares, y ello requiere un aprendizaje de corte utilitario y una idea de evaluación funcional, centrada no en lo que se puede acumular sino en lo que se sabe hacer. Lo que invita a los especialistas en evaluación y currículo a repensar el concepto de contenido implícito tanto en las pruebas nacionales como en los planes y programas de estudio de la educación básica en el país y a elaborar pruebas que demanden la aplicación de conocimientos básicos y no la repetición de un sinnúmero de temas curriculares.

Como se puede apreciar, lo importante ahora, en el plano internacional, no es medir, con base en la resolución de ejercicios descontextualizados, lo que se sabe o lo que se puede retener, sino valorar las capacidades de los estudiantes para responder a situaciones cotidianas. Ello hace necesario que en México se trabaje cada vez más en la elaboración de exámenes con reactivos que reflejen situaciones reales.

Los exámenes internacionales incluyen en el diseño de sus ítems tareas referentes a problemas de la vida diaria, que requieren la activación de procesos cognitivos de nivel superior como el razonamiento, la transferencia y la reflexión, para la aplicación e integración del conocimiento adquirido. De ahí la necesidad de incluir en los reactivos de las pruebas mexicanas tareas más de corte cognitivo que de mecanización y memoria. En adición, las pruebas internacionales incorporan modelos psicométricos recientes como el de la teoría de respuesta al ítem (TRI), en donde los ítems se seleccionan no por el número de estudiantes que los contestaron correctamente según los planteamientos de la teoría clásica del test, sino con base en una

jerarquización lógica, conceptual y cognitiva de sus propiedades que ayuda a identificar niveles de desempeño.

Los exámenes a gran escala adoptan la evaluación por niveles de desempeño y transforman los sistemas de calificación de una prueba al no limitarse al lector óptico y elaborar guías y códigos para calificar preguntas abiertas, ante lo cual es importante que las pruebas elaboradas en México por los institutos de evaluación contemplen las alternativas teórico-metodológicas que existen en materia de elaboración y calificación de exámenes y no se limiten a los procedimientos clásicos.

Al parecer, en el país no existen acciones encaminadas a la toma de decisiones pedagógicas que realmente permitan utilizar la potencialidad que dan los resultados de las pruebas para mejorar el trabajo en el aula. Por ejemplo, en materia de docencia los conceptos implícitos en los exámenes internacionales plantean la necesidad de encontrar formas de enseñanza más vivenciales, centradas en la experiencia y aplicación de un conocimiento, mediante el ejercicio y desarrollo de operaciones mentales de orden superior.

En suma, las comparaciones establecidas, permiten apreciar que los resultados obtenidos por México en estos estudios evaluativos de corte internacional como PISA, obedecen a diferencias significativas en los diversos puntos que se detallan a continuación.

Conceptualización sobre el contenido y tipo de ítems en exámenes internacionales. La comparación entre lo que se evalúa y cómo se evalúa en los ámbitos internacional y nacional revela que mientras el contenido de las pruebas internacionales gira en torno a la evaluación de *competencias*, entendidas como la aplicación del conocimiento adquirido, mediante la activación de operaciones mentales de orden superior, la mayoría de las pruebas nacionales se enfoca a la mera reproducción de información. Operaciones como el razonamiento, la interpretación y la transferencia, implícitas en las tareas de evaluación de los estudios internacionales, no son ejercitadas en las pruebas mexicanas, en donde la memoria y el recuerdo para identificar la respuesta correcta es lo fundamental. Los ítems de las pruebas internacionales en los que los estudiantes tienen que construir y justificar su respuesta son inexistentes en nuestros exá-

menes masivos, compuestos en su totalidad por ítems de opción múltiple. La construcción y selección de reactivos con base en modelos psicométricos que consideran las diferencias individuales en el desempeño de los estudiantes en términos de probabilidad y las diferentes propiedades de los ítem, no es considerada en las experiencias nacionales de evaluación, en donde todavía se recurre a un modelo psicométrico clásico, sin una definición clara y propia de sus elementos; lo que impide la detección de las fortalezas y limitaciones de capacidades cognitivas de los estudiantes.

Formas de enseñanza y evaluación. Las pruebas internacionales evalúan un aprendizaje de conceptos, hechos, principios y procedimientos, pero aplicados a escenarios reales; lo que requiere de formas de enseñanza centradas en el desarrollo de capacidades mentales que enseñen al alumno no sólo a adquirir un conocimiento sino a reflexionar sobre éste y a transferir lo que va aprendiendo a situaciones prácticas. Los estudiantes mexicanos están familiarizados con formas de enseñanza centradas en la acumulación y transmisión de información y en exámenes que miden la capacidad para retener y reproducir esa información.

Enfoques curriculares. El contenido y estructura de las pruebas internacionales requieren contenidos curriculares de tipo conceptual, pero, sobre todo, procedimental y cognitivo. Los planes y programas de estudio para la educación básica, en el Sistema Educativo Mexicano, señalan algunas habilidades que desarrollar, pero a manera de grandes objetivos, no como contenidos y procedimientos que hay que enseñar; pese a esto, se sigue privilegiando un currículo enciclopédico y sin sentido.

Aspectos socioculturales. Muchas de las tareas de evaluación de las pruebas internacionales implican un tipo de práctica sociocultural un tanto alejado de la cultura nacional. Los jóvenes en México no tienen incorporado, como parte de sus prácticas culturales, el leer sin un fin obligatorio y el reflexionar sobre el contenido y forma de lo que leen, para tomar postura y hacer uso de la lectura con diferentes fines no sólo escolares, aplicándola en situaciones cotidianas (como en la lectura de una garantía o de un esquema). Los estudiantes mexicanos no están acostumbrados a pensar matemáti-

camente o, como declaran las pruebas internacionales, a “matematizar el mundo” mediante la construcción y explicación de modelos. Las matemáticas son asumidas como algo complejo y generador de angustia, que se resuelve cuando se domina el mecanismo. El conocimiento científico es considerado como algo abstracto y sin sentido, reservado para “sabios” o gente especializada, no como una explicación de lo que ocurre en la persona y en el medio en el que se vive.

Así, las aproximaciones generadas permiten comprender el porqué de los resultados obtenidos y enunciar algunas aportaciones:

- El contenido y estructura de los exámenes internacionales llevan a pensar en la necesidad de enseñar y evaluar no sólo contenidos, sino procesos y procedimientos que favorezcan el uso del conocimiento adquirido en la resolución de los problemas de la vida cotidiana. Es decir, una tarea importante en términos de contenidos curriculares reside en enseñar no sólo lo estipulado en el plan de estudios, sino su aplicación y transferencia.
- Con miras a darle un sentido más pedagógico a la evaluación, es importante no quedarse con los resultados obtenidos, sino reparar y analizar el contenido de un examen y los conceptos de contenido, aprendizaje y evaluación implícitos en éste.
- Las funciones formativas de la evaluación sólo se podrán realizar cuando los resultados no queden en el número, calificación o puntaje alcanzado; es necesario pasar al terreno de la comprensión y analizar el por qué de los resultados obtenidos, para darle un sentido formativo y no sólo informativo a la evaluación.
- Es importante complementar y diversificar la elaboración y análisis de los instrumentos de evaluación e incorporar cuestionarios que permitan realizar estudios de contexto, para conocer no meros resultados, sino procesos y significaciones.
- La evaluación del aprendizaje debe ser abordada con un enfoque teórico y metodológico serio, y no verse reducida al plano técnico centrado en la medición.

CONSIDERACIONES FINALES

Los resultados de la prueba PISA podrían, así, ser utilizados para pensar en el tipo de educación que deseamos brindar en el sistema educativo mexicano: una enciclopédica o una que permita utilizar la información para resolver problemas. Ello permitiría analizar, además, las prácticas pedagógicas que tiene el sistema educativo desde la educación básica: queremos que los alumnos recuerden y retengan memorísticamente la información o buscamos que los estudiantes desarrollen habilidades y estrategias para usar la información y, en su caso, apropiarse de nueva información. Los resultados invitan a una revisión de los contenidos de los planes de estudio, pero también esta prueba lleva a pensar en el conjunto de condiciones socioeconómicas en el cual se desempeña la educación.

Podemos decir que la prueba PISA no ha sido “percibida en su complejidad y alcances”, pues no se ha reflexionado en torno a ciertas variables que intervienen de manera directa en este tipo de resultados. Por ejemplo, quién y cómo decidir que a los 15 años un joven o adolescente debe estar preparado para el mundo, cuando lo que debiera estar en la palestra es si “el mundo” se ha preocupado por mantenerse pendiente de niños y jóvenes desde que nacen hasta que cumplen la edad definida por la prueba.

El argumento que la OCDE enuncia sobre los 15 años es que es la edad en la cual los sistemas educativos dejan de ser obligatorios, es decir, un asunto técnico, no un argumento teórico sobre el desarrollo del individuo y la maduración de sus procesos de pensamiento.

Además, si se toma en cuenta que la escuela no es el único espacio donde un estudiante adquiere formación, sino que ésta depende también de otras esferas de la vida social como la familia, los amigos y la comunidad, entonces queda claro que dicha formación no es el resultado solamente de lo que ocurre en la escuela, como se quiere ver.

Es muy injusto pensar que una institución como la escuela pueda resolver la preparación para la vida de un sujeto, es injusto no aceptar que la escuela hoy compite con otras instancias que también formar o deforman, como los medios de comunicación o las nuevas

tecnologías aplicadas a la vida diaria. Por ello, se requiere realizar un juicio del tamaño del que genera la prueba PISA: si la escuela y su efectividad no resultan pertinentes ni adecuadas, ello no puede aislarse de los elementos contextuales presentes en la actividad.

Asimismo, cuando se establecen competencias generales y globales como las que se definen en una prueba de esta naturaleza, es crucial preguntarse si éstas deben ser compartidas por circunstancias y realidades diversas como las que puede haber entre China y Colombia o, para usar el ejemplo que ha circulado en la prensa, Finlandia y México.

Y una última cuestión: ¿los jóvenes de 15 años en estas dos realidades, deben tener las mismas competencias para la vida, aun cuando vivan y se desarrollen en contextos diferentes? Lo que sí es un hecho, porque así queda de manifiesto, es que una prueba como PISA es un buen ejemplo de orientación de políticas desde una perspectiva central, pero en ningún momento se atiende a las particularidades locales.

Desde esa perspectiva es pertinente aclarar que hay una aceptación sobre las condiciones de desarrollo económico de los países participantes en la prueba cuando se involucra en los análisis el índice de Estatus Económico, Social y Cultural (ESCS), que es finalmente el centro del tema. A partir de aquí surge una verdad por demás evidente: los países con mayor nivel de ESCS obtienen mejores resultados que aquellos que no logran altos indicadores. Luego entonces, las condiciones de desarrollo de los países determinan el desempeño de sus estudiantes: factores como el financiamiento educativo, la calidad de sus instituciones, las condiciones culturales y sociales de la población determinan los logros de los estudiantes de 15 años que responden la prueba.

De esta manera, el bajo resultado de “nuestros estudiantes no necesariamente es una responsabilidad de su esfuerzo individual, ni tampoco de la poca capacidad de sus maestros, es el resultado de vivir en un país que no logra ofrecerles condiciones de vida y desarrollo, ni a ellos, ni a sus maestros, ni a su familia”.³⁷ Así de tajante es el veredicto.

37 S. Martinic, “Información, participación, enfoque y derechos”, p. 28.

En el fondo, deberíamos felicitar a los estudiantes que hicieron el esfuerzo por responder una prueba como PISA, aunque sus condiciones escolares y sociales eran adversas y en total desventaja en comparación, digamos, con los estudiantes de Alemania, Francia o Canadá. Pero, no obstante las diferencias, los responsables de las políticas educativas de los países participantes aceptan utilizar los resultados del programa para “obtener una perspectiva externa” y orientar sus decisiones en políticas educativas.

Aunque el problema, en el caso de México, es que, a pesar de existir evidencias suficientes y ejercicios de investigación serios en el ámbito nacional, pareciera que las autoridades educativas sólo conceden importancia a los resultados de las pruebas para decidir, en el mejor de los casos, algún curso de acción.

En nuestro país no es posible negar una tradición profesional en el ámbito de la investigación educativa, pues existen instituciones y grupos que desarrollan proyectos sobre un amplio abanico de temas educativos y se da una palpable difusión de estos resultados en diferentes medios.

La propia Secretaría de Educación Pública (SEP) tiene unidades de investigación y monitoreo del sistema educativo y existe el INEE, sin olvidar que en muchas universidades existen grupos de investigación sobre educación y posgrados que desarrollan proyectos en temas educativos.

Es decir, existe un conocimiento acerca de la situación educativa en nuestro país; existe un conocimiento sobre cuáles son sus principales problemas; existe una serie de perspectivas para orientar proyectos y buscar soluciones, entonces, ¿por qué acudir a un conocimiento externo?; ¿por qué creer que los resultados de una prueba de la naturaleza de PISA pueden ofrecer lo que los resultados de la investigación local no ha enunciado?